

TEKNILLINEN KORKEAKOULU
Systeemianalyysin laboratorio
Mat-2.177 Operaatiotutkimuksen projektityöseminaari

Bioreaktorin toiminnan optimointi

Loppuraportti

14.5.2007

Petri Holappa

Tuomas Kervinen

Sampsa Kosonen

Lauri Kovanen (*projektipäällikkö*)

Ismo Räisänen

Tiivistelmä

Työssä tutkittiin vasta-aineen tuotantoon käytettävän bioreaktorin toimintaa, tarkoituksena määrittää ohjauksen kannalta kriittisiä tekijöitä vasta-aineen tuoton maksimoimiseksi. Tutkittu aineisto oli reaktorista mitattujen noin seitsemän muuttujan aikasarjatietoa, jota oli kerätty noin 70 tuotantoajosta. Ongelmaa lähestyttiin useasta suunnasta käyttäen aikasarjojen kuvaajien ja ristikorrelaatioiden silmämääräistä tarkastelua, lineaarista regressiota sopivasti varioiden sekä neuroverkkomalleja. Datassa olleiden reikien suuresta suhteellisesta määrästä sekä yleisesti ajojen lyhyestä pituudesta johtuen tulokset jäivät kuitenkin melko laihoiksi. Työtä mutkisti osin myös projektiryhmän asiantuntemuksen puute bioprosesseista. Varsinaista reaktorin dynamiikan kuvausta ei saatu aikaan, mutta varovaisesti tulkiten saaduista tuloksista voidaan tehdä johtopäätöksiä tiettyjen muuttujien merkityksestä vasta-aineen tuotannolle. Erityisesti pH-arvo sekä hapen osapaine näyttäisivät tutkimuksen perusteella nousevan merkittävimmiä vasta-aineen tuottotasoja määrittäviksi tekijöiksi. Näiden havaintojen lisäksi tutkimuksen tuloksena muodostettiin stokastinen differenssimalli, jolla voidaan simuloida eri muuttujien vaikutusta reaktorin vasta-ainepitoisuudelle, sekä ennustaa sen kehittymistä jo aloitettujen ajojen osalta. Tuloksia ei kuitenkaan ehditty validoida esimerkiksi reikien täyttämiseen käytetyn menetelmän osalta, joten niihin tulee suhtautua varauksella. Käyttökelpoisten ohjeiden tuottaminen prosessin ohjaamiseksi vaatisi tarkemmin suunnattua jatkotutkimusta nyt saatujen tulosten pohjalta.

Sisällysluettelo

Tiivistelmä	1
Sisällysluettelo.....	2
1. Johdanto	3
1.1. Tausta.....	3
1.2. Projektiryhmä ja asiakas	3
1.3. Projektityön tavoitteet	4
1.4. Bioreaktorin toiminnan kuvaus.....	4
1.4.1. Yleistä vasta-aineen tuotannosta.....	4
1.4.2. Reaktorin toimintaperiaate	4
1.4.3. Mittaukset.....	5
1.4.4. Säädot.....	6
1.5. Mallintamisessa käytetty data.....	6
1.5.1. Mitatut muuttujat ja ajojen pituus.....	6
1.5.2. Ajojen vertailukelpoisuus.....	7
2. Kirjallisuuskatsaus.....	7
2.1. Yleistä bioprosesseista	8
2.2. Lähestymistavat	8
2.3. Tulokset.....	9
3. Datan esikäsittely.....	9
3.1. Yleistä	9
3.2. Ajojen arvottaminen.....	10
3.3. Aikasarjojen reikien täyttäminen	13
4. Analysointi	15
4.1. Yleistä	15
4.2. Ajojen silmämääräinen tarkastelu.....	15
4.2.1. Menettelytapa	16
4.2.2. Havainnot	16
4.3. Tarkempi analyysi	18
4.3.1. Muuttujien ristikorrelaatiot.....	18
4.3.2. Samankaltaisten ajojen poikkeavuudet	20
4.4. Yleinen lineaarinen malli	23
4.4.1. Johdanto	23
4.4.2. Regressiomallin standardioletukset	24
4.4.3. Käyräviivaisten riippuvuuksien esittäminen	24
4.4.4. Aikasarja-aineistojen käyttöön liittyviä ongelmia	25
4.4.5. Regressiomallin estimointi datasta	26
4.4.6. Mallinrakennus regressiokertoimien estimaattien vertailemiseksi	26
4.4.7. Ennustemallin kehittäminen eri laitteilla tehdyille tuotantoajoille	33
4.4.8. Regressioanalyysi ajojen erotuksille.....	37
4.4.9. Regressioanalyysi muuttujien muutoksille.....	40
4.4.10. Pohdintaa.....	40
4.5. Bioreaktorin mallintaminen neuroverkolla.....	40
4.5.1. Neuroverkko.....	41
4.5.2. Verkon muodostaminen	42
4.5.3. Opetus ja simulointi	44
5. Johtopäätökset	50
5.1. Havainnot järjestelmän dynamiikasta	50
5.2. Suositukset tuotantoprosessin hoitamiseksi	51
5.3. Suositukset operatiivisen toiminnan kehittämiseksi	52
6. Viitteet.....	53
7. Liitteet	53

1. Johdanto

1.1. Tausta

Oy Medix Biochemica Ab on suomalainen bioalan yritys, joka tuottaa pääasiassa vasta-aineita teollisuuden käyttöön sekä erilaisia pikatestejä muun muassa raskauden ja tulehdusten toteamiseen. Yritys työllistää 80 henkilöä, ja sillä on toimipisteet Kauniaisissa ja Joensuussa.

Vasta-ainetuotanto tapahtuu hybridisoluilla *in vitro* bioreaktorissa. Toinen vaihtoehto vasta-ainetuotannolle olisi *in vivo*, eli vasta-ainetuotanto koe-eläinten avulla ascitekseenä. Tämän tyyppinen tuotanto on kuitenkin EU:n alueella tiukasti säädeltyä ja koko maailmassa vallitseva suuntaus on ollut vähittäinen siirtyminen kohti bioreaktoryyppistä vasta-ainetuotantoa.

Vasta-aineen tuotantoprosessia ohjataan ainoastaan muuttamalla nestepumppujen nopeutta. Pumput tuovat bioreaktoriin ravintoliuosta, ja niiden avulla kerätään myös *harvesti*, eli prosessin lopputuote, vasta-aine. Tarkempi kuvaus bioreaktorin toiminnasta on kappaleessa 1.4.

Tuotantoprosessin ongelmana on se, että vasta-aineen tuotanto saattaa vaihdella suuresti ajosta toiseen. Joskus hyvin alkanut ajo saattaa tyrehtyä ilman päällepäin näkyvää syytä, ja joskus taas tuotanto ei saavuta kannattavaa tasoa lainkaan. Vasta-ainetuotannon haasteena onkin se, että prosessin tuottoa on hyvin vaikea ennakoida ja lisäksi ei tiedetä ohjausta, jolla tuotannon saisi pidettyä mahdollisimman lähellä optimia. Tuotantoprosessin hoidon ammattitaito on kyllä kehittynyt sitä kautta, että sen parissa toimivat työntekijät ovat hoitaneet tehtävänsä hyvin kauan. Tämä ammattitaito ei kuitenkaan riitä tuottamaan riittävän perusteltuja johtopäätöksiä reaktorin dynamiikasta ja lisäksi häviää talosta työntekijöiden siirtyessä eläkkeelle.

Projekti toteutettiin kurssiin Mat-2.177 Operaatiotutkimuksen projektityöseminaari kuuluvana ryhmätyönä kevään 2007 aikana. Projektin tavoitteet on esitelty tarkemmin kappaleessa 1.3, sekä projektiin osallistuneet henkilöt ja heidän roolinsa kappaleessa 1.2.

1.2. Projektiryhmä ja asiakas

Projektiryhmä koostui systeemi- ja operaatiotutkimusta pää- tai sivuaineenaan lukevista opiskelijoista. Projektiryhmän jäsenet olivat **Petri Holappa**, **Tuomas Kervinen**, **Sampsa Kosonen**, **Lauri Kovanen** (projektipäällikkö) ja **Ismo Räisänen**.

Projektin asiakas oli **Oy Medix Biochemica Ab** ja asiakkaan yhteyshenkilönä toimi tutkimus- ja tuotekehityspäällikkö **Matti Höyhty**. Asiakkaan roolina projektissa oli taustaprosessin tuntemuksen antaminen mallinnuksen tueksi ja erityisen tärkeänä myös sen valvonta, että projektin tavoitteet on asetettu oikein ja projekti etenee niitä kohti.

Lisäksi projektin parissa toimi kurssihenkilökunta, johon kuuluivat professori **Ahti Salo** ja assistentti **Jussi Kangaspunta**. Kurssihenkilökunnan kanssa käytiin keskustelua projektin käytännön etenemisestä ja tutkimustyön toteuttamistavasta. Lisäksi kurssihenkilökunta järjesti kurssiin kuuluvat tilaisuudet (3

kpl), joissa esiteltiin projektisuunnitelma, sen eteneminen sekä tämä loppudokumentaatio.

1.3. Projektityön tavoitteet

Projektin tavoitteiksi asetettiin:

- Muodostaa reaktorin toimintaa kuvaava malli, selitettävänä muuttujana vasta-aineen tuotanto tietyllä ajanhetkellä
- Antaa suositukset tuotantoprosessin hoitamiseksi jatkossa mallin pohjalta
- Harjoitella suunnitelmallista projektityöskentelyä

Lähtökohtaisesti reaktorin toimintaa kuvaavasta mallista pyrittiin tekemään mahdollisimman yksityiskohtainen ja selittävä; ihannetilanne olisi yksi malli, jolla pystyttäisiin kuvaamaan kaikkia odotettavissa olevia tapauksia, ja joka ennustaisi riittävällä tarkkuudella sellaiset tuotantoajat, joita ei käytetä mallin parametrien estimoinnissa. Toisaalta yksinkertaisempi lopputulos on kausaalityyppinen malli, joka kuvaisi karkealla tasolla parametrien ja tuotantoajon onnistumisen riippuvuuden.

Jo aluksi tiedostettiin, että annettavat suositukset tuotantoprosessin hoitamiseksi riippuvat vahvasti muodostetusta mallista. Parhaassa tapauksessa mallista voitaisiin ratkaista *optimaalinen ohjaus*, kun taas toisessa ääripäässä jouduttaisiin tyytymään perusteltuihin suosituksiin tarkkailtavien muuttujien *hälytysrajoista*.

Suunnitelmallisen projektityöskentelyn harjoittelu kuului yhtenä kurssilla opetettaviin asioihin.

1.4. Bioreaktorin toiminnan kuvaus

1.4.1. Yleistä vasta-aineen tuotannosta

Vasta-ainetta tuottavat solut on tuotettu yhdistämällä syöpäsoluja ja lymfosyyttejä eli imusoluja. Tuloksena saadaan hybridisolun, joka jatkaa jakautumista käytännössä ikuisesti. *Solulinja* on yhdestä tällaisesta solusta monistamalla saatu solujoukko, jota voidaan varastoida, ja josta otetaan aina pieni määrä soluja yhtä tuotantoajoa varten.

Jotta solut saadaan tuottamaan vasta-ainetta, tarvitaan sopiva kasvatusalusta. Tämä voi olla elävä eläin, yleisimmin hiiri tai vaihtoehtoisesti eloton alusta. Tässä työssä tutkittu bioreaktori on yksi muoto jälkimmäisistä.

Tutkitussa tuotantoprosessissa tuotettu vasta-aine oli hiiren immunoglobuliini (IgG)

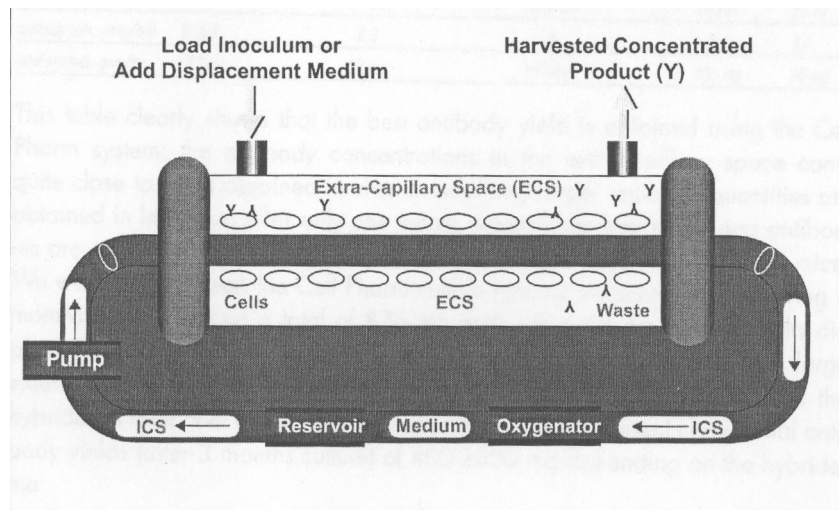
1.4.2. Reaktorin toimintaperiaate

Bioreaktori sisältää kapillaariputkia, joiden väliseen tilaan vasta-ainetta tuottavat solut injektoidaan ennen tuotantoajon aloittamista. Putkien seinämät ovat puoliläpäisevää kalvoa, joka päästää lävitseen pienimolekyyliset aineet, kuten kaasut ja prosessin ravintoaineet, muttei itse soluja tai niiden tuottamaa vasta-ainetta.

Kapillaariputkien sisälle johdetaan ravintoliuosta jatkuvana virtana. Ravintoliuoksesta kulkeutuu diffuusion avulla solutilaan ravintoaineita ja päinvastoin solutilasta poistuu metaboliatuotteita, kuten laktaattia ja ammoniumioneja.

Solutilassa olevaa nestettä kierrätetään erillisellä pumpulla, joka pumppaa nesteen harvestointiastiaan. Tämä lopputuotteen (= vasta-aine) sisältävä liuos edelleen konsentroidaan varastointia, puhdistamista ja asiakkaalle toimittamista varten.

Bioreaktorin toimintaa on havainnollistettu oheisessa kuvassa (Kuva 1).



Kuva 1: Bioreaktorin toimintaperiaate.

1.4.3. Mittaukset

Bioreaktorista mitataan useita muuttujia, joista mallinnuksen kannalta olennaisimmat ovat:

- Hapen osapaine (mg/l)
- Glukoosipitoisuus (mg/dl)
- Laktaattipitoisuus (mg/dl)
- pH-arvo
- Solutilan vasta-ainepitoisuus (mg/ml)

Tämän lisäksi harvestointien yhteydessä mitattiin seuraavia muuttujia:

- Harvestoidun liuoksen vasta-ainepitoisuus (mg/ml)
- Harvestoidun liuoksen tilavuus (ml)
- Prosessin kumulatiivinen vasta-aineen tuotanto (mg)

Lisäksi pumppujen kiertonopeudet tiedetään jokaisella hetkellä riippumatta muista mittauksista, koska niitä säädetään käsin.

Aineiden pitoisuudet vasta-ainetta lukuun ottamatta mitataan ravintoliuoksen ulostulovirtauksesta.

Solutilassa olevat pitoisuudet lasketaan sen erotuksen avulla, joka tietyn aineen pitoisuudessa on solutilaan menevän ja sieltä tulevan virtauksen välillä. Vasta-ainepitoisuus mitataan harvestista.

Näiden muuttujien lisäksi jokaisesta ajosta on tiedossa ajon alussa kirjattu soluluku ja solujen elinvoima. Soluluku kertoo sen, montako solua reaktoriin on laitettu ja elinvoima määrittää elävien solujen prosenttiosuuden.

1.4.4. Säädot

Tuotantoprosessia voidaan säätää ainoastaan muuttamalla pumppujen kiertonopeutta.

- Ravintoliuosta kuljettavalla sirkulaatiopumpulla vaikutetaan suoraan erityisesti glukoosi- ja laktaattipitoisuuksiin. Mediapumpulla syötetään tuoretta ravintoliuosta tähän sirkulaatiokiertoonsä sekä poistetaan metaboliatuotteita kierrosta.
- Harvestointipumpulla kerätään solutilassa olevaa nestettä (= harvesti), johon solut erittävät vasta-aineen (= lopputuote). Harvestointipumppu ei poista soluja, vaan ainoastaan vasta-ainetta.

Näiden lisäksi bioreaktori säätää automaattisesti prosessin happamuutta (eli pH:ta) happi-hiilidioksidi -suhdetta muuttamalla.

1.5. Mallintamisessa käytetty data

1.5.1. Mitatut muuttujat ja ajojen pituus

Mallintamisessa käytettiin aineistoa, johon kuului yhteensä 69 tuotantoajoa. Ajoista 33 saatiin valmiiksi sähköisessä muodossa, ja 36 ajoa kerättiin Excel-taulukoihin käsin paperille kirjatusta tiedosta. Käytetyt ajot on eritelty tarkemmin kappaleessa 3.2.

Sähköisessä muodossa saaduissa ajoissa mittaustieto oli kerätty joka arkipäivä, kun taas paperilta kerätyissä ajoissa havaintoja oli vain noin kaksi viikossa. Aineistojen sisältämät tiedot vaihtelivat myös sen mukaan, mistä muuttujista tietoa oli kerätty; kaikki ajot eivät sisältäneet tietoa kaikista ylipäänsä mitatuista muuttujista. Kaikille ajoille yhteisiä olivat vain seuraavat muuttujat:

- Glukoosi
- Laktaatti
- Vasta-ainepitoisuus
- Harvestointien yhteydessä mitatut muuttujat:
 - Harvestoinnin volyyymi
 - Harvestoidun liuoksen vasta-ainepitoisuus
 - Kumulatiivinen vasta-ainetuotto
- Pumppujen kiertonopeudet:

- Harvestointipumppu
- Mediapumppu
- Sirkulaatiopumppu

Harvestointeja oli tehty noin viikon välein, ja ne oli kirjattu täydellisesti kaikkien ajojen osalta.

Käytettyjen tuotantoajojen pituus vaihteli noin viikosta kuuteen kuukauteen. Hyviksi luokitelluista (ks. kappale 3.2) ajoista lyhin oli kuitenkin jo lähes kahden kuukauden mittainen, joten lyhimpiin ajoaikoihin syynä oli huono tuotto tai prosessin muu epäonnistuminen ulkoisista tekijöistä johtuen. Havaintoja hyvissä, käsin sähköiseksi koodatuissa ajoissa oli suurin piirtein 25-42 kappaletta.

1.5.2. Aojen vertailukelpoisuus

Aojen vertailukelpoisuuteen vaikuttaa kaksi tekijää: tuotantoon käytetty laite sekä solulinja.

Mallintamiseen käytetyssä datassa vasta-aineen tuotantoon oli käytetty kolmea eri laitetyyppiä, jotka eroavat toisistaan sen suhteen, kuinka monta bioreaktoria niissä on. Laitteiden sisällä olevat bioreaktorit ovat identtisiä. Laitetyypit olivat (suluissa reaktorien määrä):

- Acu R (1)
- Maximizer (2)
- Xcell FP1 ja FP2 (6)

Näiden tyyppien välillä eroja on siis vain kulutetun ravintoliuoksen sekä tuottona saadun vasta-aineen volyymissa. Volyymien voi karkeasti laskea olevan suoraan verrannollisia reaktorien määrään.

Käytetyn solulinjan vaikutus maksimaaliseen (teoreettiseen) tuottotasoon sen sijaan on epäselvempi. Käsin kerätyn aineiston ajot oli kokonaisuudessaan tehty samalla solulinjalla, joten niiden vertailu oli suoraan mahdollista; erottavana tekijänä niissä oli vain käytetty reaktori. Sähköisessä muodossa saadussa aineistossa taas oli käytetty yhteensä 24 eri solulinjaa. Nämä solulinjat on mainittu myöhemmin taulukossa (Taulukko 1).

Näiden solulinjojen välillä saavutetun tuottotason vertaaminen on vaikeaa. Sen sijaan eri muuttujien suhteellinen käyttäytyminen pitäisi olla samanlaista solulinjasta riippumatta; esimerkiksi väite "pH:n noustessa yli hälytysrajan vuorokausituotto putoaa 95 % todennäköisyydellä 50 %" voisi päteä kaikille ajoille reaktorityypistä ja käytetystä solulinjasta riippumatta.

2. Kirjallisuuskatsaus

Erilaisista bioreaktoreiden säätö- ja optimointiongelmien ratkaisuksista löytyi useita erilaisia tutkimuksia. Päädyimme kuitenkin rajaamaan aikaisempien tutkimusten lähemmän tarkastelun kahteen eri tutkimukseen, koska niiden aiheisältö oli lähinnä meidän työtä. Tutkimuksista ensimmäinen käsittelee bioreaktoreiden dynaamista optimointia [1] ja toinen fed-batch bioreaktoreiden optimointia hyödyntäen

neuroverkkoja [2].

2.1. Yleistä bioprosesseista

Bioprosessit voidaan karkeasti ryhmitellä seuraaviin kolmeen ryhmään: jatkuviin, erätuotannollisiin ja fed-batch -prosesseihin. Fed-batch -prosesseissa solut ja mikro-organismit kasvatetaan bioreaktoreissa, missä ravintoliuosta tarjotaan koko prosessin ajan ohjatulla ruokinnalla.

Fed-batch -prosesseilla on useita tunnettuja etuja verrattuna jatkuviin ja erätuotannollisiin bioreaktoreihin. Fed-batch on esimerkiksi usein paras ja jopa ainoa vaihtoehto tehokkuutensa ansiosta tapauksissa, joissa ei-toivotut ilmiöt esiintyvät kuten kasvualustan inhiboituminen ja kataboliittinen repressio. Inhiboitumisessa on kyse kemiallista reaktiota heikentävästä ilmiöstä ja kataboliittinen repressio tarkoittaa biokemiassa solujen toimintojen lakkaamista eli toisin sanoen solujen tukehtumista. Lisäksi fed-batch tarjoaa paremman poikkeamien ohjattavuuden organismien kasvumallissa.

2.2. Lähestymistavat

Aikaisemmissa tutkimuksissa oli löydettävissä muutamia bioreaktoreiden optimointimalleja, joissa on sovellettu dynaamista optimointia maksimoimaan proteiinin tuotanto. Näistä mainittakoon kaksi eri optimointimallia, joiden nimet ovat: Park-Ramirez ja Lee-Ramirez bioreaktorit. Park-Ramirez bioreaktorissa tavoitteena on maksimoida eriytynyt proteiini fed-batch viljelmästä. Lee-Ramirez bioreaktorissa on laajennettu ensin mainitun bioprosessin toimintaa. Sen tarkoituksena on maksimoida prosessin kokonaishyöty tietyssä ajassa.

Bioreaktoreiden optimointiongelmat ovat usein ratkaistu dynaamisella optimoinnilla. Dynaamisen optimoinnin käyttö edellyttää kuitenkin dynaamisen mallin olemassaoloa systeemistä. Tällainen lähestymistapa on hyvin malliriippuvainen, minkä vuoksi pienilläkin epätarkkuuksilla systeemin mallissa voi olla suuri merkitys saataviin tuloksiin. Dynaamisten mallien käyttö voi lisäksi johtaa dimensionaalisuuden kiroukseen.

Dynaamisten mallien numeeriset ratkaisut on tavallisesti jaettu kolmeen kategoriaan: epäsuoriin, suoriin lähestymisiin sekä dynaamiseen ohjelmointiin. Epäsuorat (klassiset) lähestymistavat ovat useiden tutkijoiden käyttämiä. Niissä käytetään variaatiolaskennan tekniikoita optimoimaan bioreaktori. Epäsuoria lähestymistapoja voidaan käyttää erottamaan ratkaisun yleisiä ominaisuuksia toisistaan.

Suorat menetelmät ovat nykyään yleistyneet ratkaistaessa dynaamisia optimointiongelmia. Suorat menetelmät on jaettu kahteen eri strategiaan: ohjausvektorin parametrisointiin ja täydelliseen parametrisointiin. Ohjausvektorin parametrisoinnissa ainoastaan ohjausmuuttujat on parametrisoitu, kun taas täydellisessä parametrisoinnissa sekä ohjaus- että tilamuuttujat on parametrisoitu.

Useimmat bioteknologian reaktorit ovat epälineaarisia ja niillä on ajasta riippuva dynamiikka. Siten kokeiden toistettavuus on suhteellisen heikkoa. Lisäksi bioprosessien vasteet ovat keskimäärin hyvin hitaita. Bioreaktorit ovat useimmiten fed-batch tyyppisiä. Siten systeemien dynamiikka tulee usein

äärimmäisen monimutkaiseksi. Bioprosessien dynamiikan ollessa epälineaarinen, tila- ja ohjausmuuttujiin tulee rajoituksia, jotka on syytä huomioida mallinnettaessa bioprosesseja.

Yleisesti ottaen bioreaktorisysteemit ovat vaikeasti mallinnettavissa olevia prosesseja. Usein prosessin mekanismit ja niiden järjestykset ovat tuntemattomia ja ovat usein vain kokeellisesti vahvistettuja. Neuroverkot tarjoavat usein tällaisissa tapauksissa ongelmaan kohtuullisen hyvän ratkaisun. Neuroverkkosysteemi opettelee systeemin dynamiikan matkimalla todellisen prosessin kulkua hyödyntäen prosessihistoriaa ennustuksien laatimisiin. Neuroverkkosysteemillä voidaan saada aikaan kokeellinen malli, jonka avulla bioreaktorin mahdollinen optimointi voidaan myöhemmin sitten suorittaa ottamalla huomioon tiedetyt tasapainosuhteet.

2.3. Tulokset

Bioreaktoreiden dynaamisessa optimoinnissa [1] oli kiinnitetty erityistä huomiota tapaan, miten ohjausvektorit parametrisoidaan. Ohjausvektoreiden parametrisointi on suora menetelmä, joka kääntää alkuperäisen ongelman epälineaariseksi ongelmaksi. Tämä epälineaarinen ongelma ratkaistaan sitten sopivaa menetelmää käyttäen. Systeemien epälineaarisuuksista, erilaisista rajoituksista ja epäjatkuvuuksista johtuen, numeeriset ongelmat nousivat kuitenkin huomattavan suuriksi.

Neuroverkoissa [2] päädyttiin käyttämään systeemin dynamiikan mallintamisessa mieluummin konservatiivisia helposti saatavilla olevia peruseriaatteita prosessien dynamiikasta kuin black-box-malleja. Neuroverkkojen soveltamista ja dynaamisen ohjelmoinnin soveltuvuutta testattiin Park-Ramirez- ja Lee-Ramirez-bioreaktoreilla. Approksimaatiovirheet neuroverkoissa ja täydellisten tilainformaatioiden puuttuminen johtivat osaoptimointiin. Kuitenkin tuloksien perusteella voitiin päätellä, että neuroverkot tarjoavat nopean ja luotettavan menetelmän bioreaktorien mallintamiseen ja optimointiin.

3. Datan esikäsittely

3.1. Yleistä

Mallinnuksen aloittaminen edellytti, että saadulle aineistolle tehtiin projektin työmäärään nähden melko suuri esikäsittely. Esikäsittelyyn kuuluivat datassa olevien reikien täyttäminen sekä ajojen arvottaminen hyvyyden mukaan.

Reikien täyttö oli välttämätön toimenpide tilastollisen analyysin mahdollistamiseksi. Vaikka eri muuttujien havainnot olivatkin aina samoilta päiviltä, mittauspäivien väliajoissa oli eroja. Näin mittauksia ei voitu vain yhdistää peräkkäisiksi havainnoiksi, vaan tarvittiin menetelmä mittauspisteiden välisten "havaintojen" generoimiseksi.

Tämän lisäksi mallintamista varten tarvittiin selitettävä tekijä. Alunperin ajeltiin, että jokaiselle ajolle muodostettaisiin ajon hyvyyttä kuvaava yksikäsitteinen muuttuja, jota selitettäisiin muiden muuttujien vaihtelulla. Tämä suunnitelma muuttui hieman työn edetessä, mutta vertailukelpoisten ajojen keskinäisen

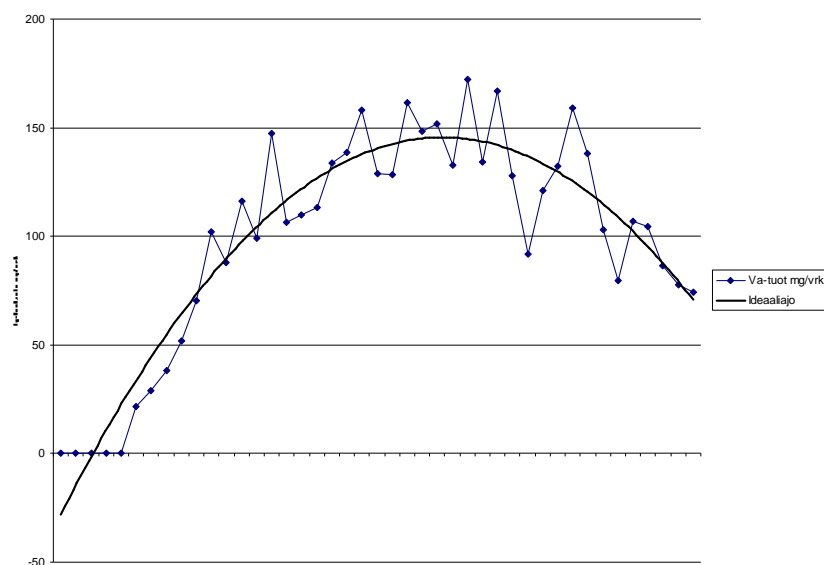
arvojärjestyksen määrittäminen katsottiin kuitenkin tarpeelliseksi.

Esikäsittelyyn käytetyt menetelmät on kuvattu tarkemmin seuraavissa kappaleissa.

3.2. Ajojen arvottaminen

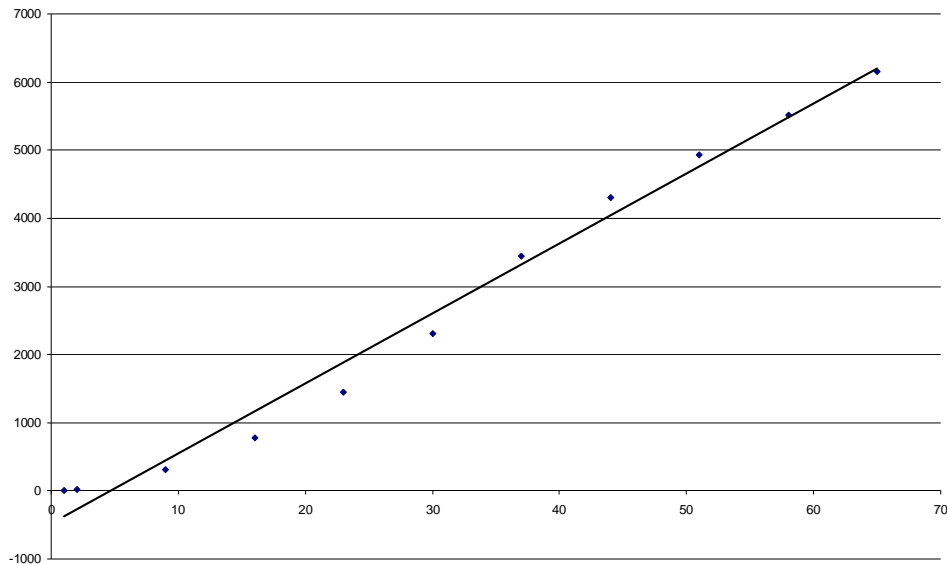
Jonkinlaisen hyvyyskriteerin muodostaminen tuotantoajoille oli välttämätöntä, jotta ajot saataisiin keskenään järjestykseen, ja niitä pystyttäisiin vertailemaan onnistumisen suhteen. Alun perin oli lisäksi ajatuksena, että mallinnuksessa selitettävänä tekijänä voisi käyttää nimenomaan mallin hyvyttä. Tuotantoajot arvotettiin ennen datassa olevien reikien täyttämistä.

Hyvyyskriteeri ajateltiin muodostaa laskemalla ensin “ideaalisen ajon” vuorokausituottoon sovitettu käyrä. Tämän jälkeen satunnaisen ajon hyvyys saataisiin laskemalla tämän ajon ja *ideaaliajon* välinen differenssi summattuna koko ajon pituudelta. Menetelmää on havainnollistettu oheisessa kuvassa (Kuva 2). Emme kuitenkaan toimineet näin, koska toistensa kanssa vertailukelpoisia ajoja oli lopulta alun perin luultua vähemmän (ks. kappale 1.5.2) ja menetelmä olisi ollut melko työläs sen suhteen, että jokaiselle reaktori- ja solulinjakombinaatiolle olisi pitänyt estimoida oma ideaaliajo. Tämän takia päädyttiinkin tarkkuudeltaan samanlaatuiseen, mutta yksinkertaisempaan ratkaisuun.



Kuva 2: Vuorokausituottoon sovitettu toisen asteen ideaaliajopolynomi ajolle HESSU2

Havaitsimme nimittäin, että kumulatiivinen tuotto noudattaa varsinkin tuotoltaan parhaimmissa ajoissa likimain suoraa. Teimme ajosten arvottamisen kumulatiiviseen tuottoon sovitetun regressiosuoran avulla. Suoran kulmakerroin kertoo näin tuottoasteen ja regression selitysaste tuoton tasaisuuden. Periaatetta on havainnollistettu allaolevassa kuvassa (Kuva 3).



Kuva 3: Ajon HESSU2 kumulatiivinen tuotto ja siihen sovitettu regressiosuora

Kun ajot arvotetaan tällä menetelmällä, myös prosessin hoitajalla on vaikutusta ajon hyvyyteen, sillä kumulatiivisen tuoton kertymiseen vaikuttaa harvestointien taajuus ja määrä. Tämän ei katsottu kuitenkaan haittaavan, sillä merkitys on melko pieni (prosessin hoitajat pyrkivät mahdollisimman tasaiseen tuottoon) ja myös alun perin suunniteltu sofistikoituneempi arvotusmenetelmä olisi ollut sille altis.

Tuotantoajot järjestettiin allaolevan taulukon (Taulukko 1) mukaiseen järjestykseen. Ajot järjestettiin laitetyypeittäin regressiosuoran kulmakertoimen mukaan ja lisäksi niiden hyvyydelle annettiin sanallinen arvio kumulatiivisen tuottokäyrän nopean tutkimisen perusteella.

Acu jr-, Maximizer- ja Xcell-ajot yksilöitiin laitetypin mukaisella nimellä ja juoksevalla numeroinnilla. Valmiiksi sähköisessä muodossa olleet vertailukelpoiset ajot nimettiin vastaavasti solulinjan mukaan. Arvottomat jääneet ajot nimettiin muista erottamiseksi nimellä AJOxxx, jossa xxx on alkuperäinen ajon numero¹. Näin jokaiselle ajolle saatiin havainnollinen ja yksilöllinen tunniste.

Taulukko 1: Tuotantoajot laitteen mukaan hyvyysjärjestyksessä

Ajon tunniste	Laitetyyppi	Solulinjan tunniste	Kulmakerroin	Selitysaste	Hyvyys	Käyttö
AKU1	Acu R	1	260,2	0,97	Hyvä, vähän havaintoja	
AKU2	Acu R	1	167,2	0,95	Hyvä	
AKU3	Acu R	1	158,6	0,97	Hyvä	V
AKU4	Acu R	1	142,9	0,99	Hyvä	
AKU5	Acu R	1	106,4	0,97	Hyvä alku	
AKU6	Acu R	1	60,8	0,97	Hyvä alku	
AKU7	Acu R	1	60,7	0,99	Kohtalainen	
AKU8	Acu R	1	56,1	0,99	Kelvoton	K
AKU9	Acu R	1	28,7	0,72	Huono, vähän havaintoja	

¹ Tätä numeroa ei ollut määritelty tai saatavilla kaikille ajoille, joten sitä ei voitu käyttää tunnisteenä. Se olisi tietysti ollut johdonmukaisin ratkaisu ajot yksilöiväksi tunnisteeiksi.

MAX1	Maximizer	1	237,5	0,99	Erittäin hyvä	
MAX2	Maximizer	1	148,7	0,99	Hyvä	
MAX3	Maximizer	1	143,0	0,98	Hyvä	
MAX4	Maximizer	1	138,8	0,94	Erittäin hyvä	V
MAX5	Maximizer	1	135,6	0,87	Hyvä, huono muoto	
MAX6	Maximizer	1	110,3	0,97	Hyvä alku	
MAX7	Maximizer	1	96,9	1,00	Kelvoton	K
MAX8	Maximizer	1	89,3	0,96	Kelvoton	K
MAX9	Maximizer	1	42,1	0,95	Huono	
MAX10	Maximizer	1	39,2	0,96	Huono	V
MAX11	Maximizer	1	12,4	1,00	Kelvoton	K
XAVIER1	Xcell	1	927,4	0,92	Erittäin hyvä	
XAVIER2	Xcell	1	834,7	0,99	Erittäin hyvä	
XAVIER3	Xcell	1	507,3	0,99	Hyvä	
XAVIER4	Xcell	1	494,7	0,96	Hyvä	V
XAVIER5	Xcell	1	485,1	0,98	Hyvä alku, notkahdus	
XAVIER6	Xcell	1	474,1	0,99	Hyvä	
XAVIER7	Xcell	1	400,2	0,98	Hyvä alku, notkahdus	
XAVIER8	Xcell	1	381,4	0,98	Hyvä	
XAVIER9	Xcell	1	357,6	0,99	Kohtalainen	
XAVIER10	Xcell	1	310,9	0,98	Kohtalainen	V
XAVIER11	Xcell	1	307,8	0,99	Kohtalainen	
XAVIER12	Xcell	1	290,4	0,98	Kohtalainen	
XAVIER13	Xcell	1	258,1	0,99	Kohtalainen	
XAVIER14	Xcell	1	186,5	1,00	Huono	V
XAVIER15	Xcell	1	142,8	0,99	Huono	
XAVIER16	Xcell	1	32,5	1,00	Kelvoton	K
PAS11	Acu R	2	140,2	0,99	Hyvä	
ADAM1	Acu R	3	124,4	0,99	Hyvä	
KERKKO1	Acu R	4	120,9	0,99	Hyvä	
AJO222	Acu R	5	105,4	0,98		EA
AJO202	Acu R	6	92,8	0,95		EA
HESSU1	Acu R	7	91,6	0,97	Hyvä	
HESSU2	Acu R	7	102,7	0,99	Hyvä	
AJO221	Acu R	8	85,5	0,76		EA
AJO232	Acu R	9	85,0	1,00		EA
AJO207	Acu R	10	83,6	0,90		EA
AJO223	Acu R	8	70,7	0,89		EA
AJO230	Acu R	11	70,0	0,74		EA
AJO233	Acu R	12	68,5	0,69		EA
AJO203	Acu R	13	66,8	0,97		EA
AJO212	Acu R	14	65,4	0,94		EA
AJO220	Acu R	5	61,6	0,80		EA
AJO213	Acu R	11	57,1	0,95		EA
AJO227	Acu R	15	49,8	0,99		EA
ANTERO1	Acu R	16	45,9	0,89	Kohtalainen	
AJO216	Acu R	17	43,7	0,97		EA
ANTERO2	Acu R	16	41,8	0,99	Kohtalainen	
AJO224	Acu R	4	40,4	0,74		EA
AJO218	Acu R	18	33,2	0,89		EA
AJO219	Acu R	19	23,2	0,98		EA

ADAM2	Acu R	3	19,7	0,94	Huono	
AJO214	Acu R	20	17,2	0,68		EA
ADAM3	Acu R	3	14,1	1,00	Huono	
ANTERO3	Acu R	16	6,7	0,94	Huono	
AJO205	Acu R	21	5,1	0,90		EA
AJO206	Acu R	22	3,7	0,88		EA
AJO215	Acu R	23	0,0	0,00		EA
AJO229	Acu R	11	0,0	0,00		EA
AJO234	Acu R	12	0,0	0,00		EA

V = validointiajo

K = käyttökelvoton

EA = ei arvioitu ajon hyvyttä

Tässä vaiheessa päätettiin myös jättää mallintamisesta pois osa ajoista muodostettavien mallien validointia varten. Lisäksi osa ajoista katsottiin käyttökelvottomiksi, koska niissä ei ollut yhtään harvestointia tai tuotantoaineistoa oli vain hyvin lyhyeltä pätkältä.

Taulukosta nähdään, että tuotantoajot jakautuvat aika tasaisesti kolmeen ryhmään: hyvät, kohtalaiset ja huonot. Xcell-laitteella tehtyjen ajojen tuotot ovat selkeästi muita suuremmat, kuten pitääkin. Keskimääräiset tuotot eivät kuitenkaan skaalaudu aivan lineaarisesti reaktorien määrän mukaan. Maximizerilla tehdyissä ajoissa tuotot eivät ole keskimäärin suurempia Acu jr -laitteeseen verrattuna, toisin kuin bioreaktorien määrästä voisi päätellä.

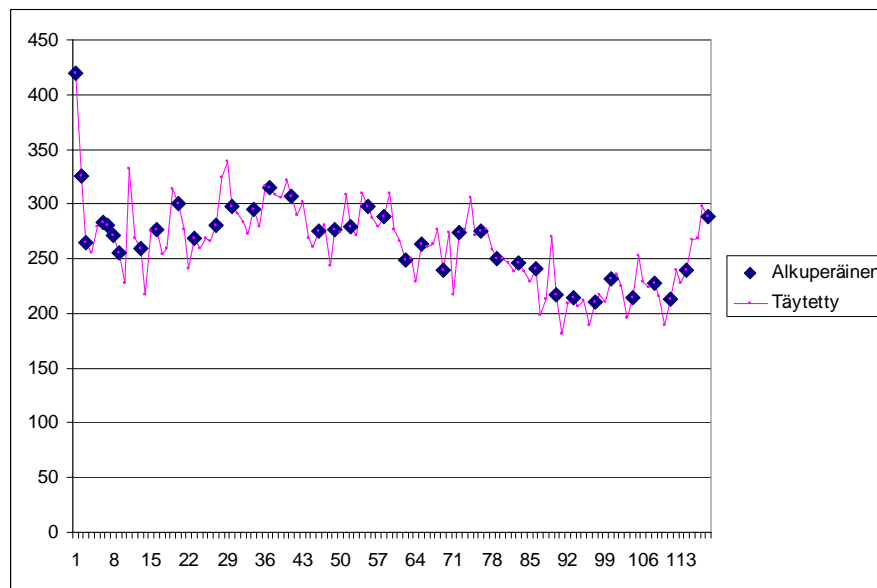
Taulukko havainnollistaa myös, kuinka kirjavaa valmiiksi sähköisessä muodossa ollut aineisto on; siitä löytyy vain muutama keskenään vertailukelpoinen ajo.

3.3. Aikasarjojen reikien täyttäminen

Koska käsin sähköiseen muotoon muutettua mittausdataa oli kerätty keskimäärin kaksi kertaa viikossa, aikasarjat olivat hyvin reikäisiä. Jotta analyysin suorittaminen olisi ollut mahdollista, aikasarjat piti saada täytettyä. Koska kaikkien muuttujien havainnot olivat useimmiten samoilta päiviltä, regressiomalleihin perustuvia täyttämismenetelmiä ei voitu käyttää. Ne nimittäin vaativat, että datan reiät ovat satunnaisia, kun tässä tapauksessa ne olivat säännöllisiä. Koska projektin aiheena oli tutkia bioreaktorin toimintaan vaikuttavia tekijöitä, täyttämiseen ei haluttu käyttää paljoa aikaa. Niinpä hienostuneet menetelmät, kuten Markovin ketjuihin ja Monte Carlo -simulointiin perustuvat menetelmät hylättiin ja päädyttiin lineaariseen interpolaatioon perustuvaan täyttömenetelmään. Täyttämisen yhteydessä lisättiin hieman satunnaisvaihtelua arpomalla normaalijakaumasta satunnaisen häiriön. Jokaiselle muuttujalle arvottiin aina uusi satunnainen häiriö, mikä pienentää muuttujien välisiä riippuvuuksia. Pienenemisen oletettiin kuitenkin olevan niin pieni, että sen ei pitäisi aiheuttaa suuria vaikeuksia.

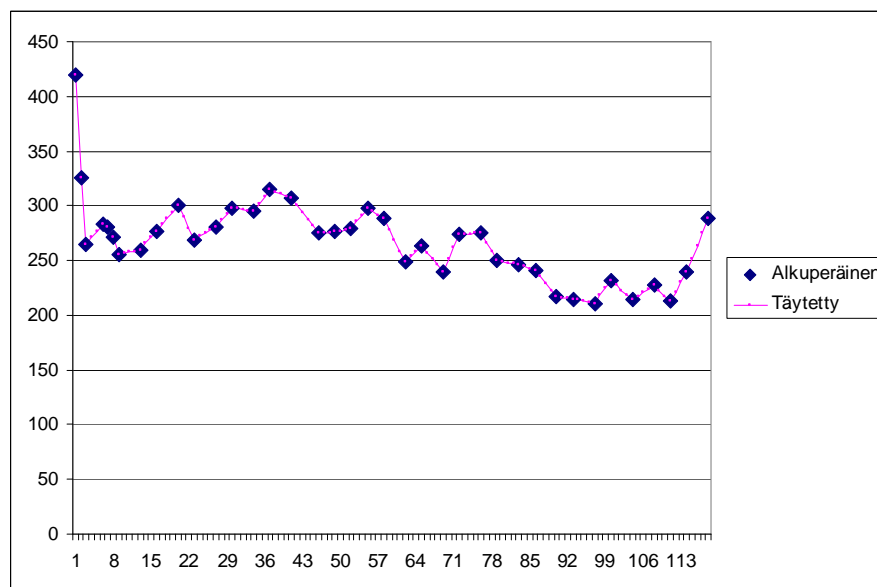
Satunnaisen häiriön odotusarvon oletettiin olevan nolla ja otoskeskihajonta määritettiin seuraavasti. Ensin laskettiin otoskeskihajonta käyttäen alkuperäisen, täyttämättömän aikasarjan havaintoja, jotka olivat seitsemältä edelliseltä ja jälkeiseltä päivältä täyttökohtaan nähden. Tämän jälkeen saatuja otoskeskihajontoja tasoitettiin ottamalla 14 päivän liukuva keskiarvo. Tasoitus sen vuoksi, että lasketut

otoskeskihajonnat vaihtelivat melko paljon. Esimerkki täytetystä aikasarjasta on esitetty allaolevassa kuvassa (Kuva 4).



Kuva 4: Esimerkki täytetystä aikasarjasta. Satunnaisvaihtelun lisääminen aiheuttaa sen, että aikasarja saattaa vaihdella melko paljon alkuperäisten pisteiden välillä.

Jälkikäteen saatujen tietojen perusteella, satunnaishäiriön lisääminen aikasarjojen täyttämässä on tarpeetonta. Olisi kannattavampaa käyttää pelkkää lineaarista interpolaatiota. Tällöin täytyy kuitenkin ottaa huomioon, että estimoinneissa saadut luottamusvälit voivat olla liian optimistisia. Todellisuudessa välit ovat todennäköisesti leveämpiä. Linearisella interpolaatiolla täytettyjä aikasarjoja käytettiin regressioanalyysissä, kun tutkittiin ajosten muuttujien välisten erojen vaikutuksia. Esimerkki lineaarisella interpolaatiolla täytetystä aikasarjasta on oheisessa kuvassa (Kuva 5).



Kuva 5: Esimerkki lineaarisella interpolaatiolla täytetystä aikasarjasta.

4. Analysointi

4.1. Yleistä

Aineiston esikäsittelyn jälkeen oli itse mallintamisen² vuoro.

Kirjallisuusselvityksen jälkeen vaikutti siltä, että käytettävissä oleva aika ja projektiryhmän asiantuntemus eivät riittäneet siihen, että olisimme muiden näyttämien suuntaviivojen mukaisesti lähteneet suoraan tavoittelemaan järjestelmän toiminnan täysin kuvaavaa tilayhtälömallia. Toisaalta myöskään asiakkaan tarpeissa ei ollut vasta-aineen tuotantoprosessin dynamiikan täydellinen kuvaus, vaan prosessin hoitotavasta johtuen suurin hyöty olisi nimenomaan selkeillä, toteutettavissa olevilla ohjenuorilla, jotka keskimäärin parantaisivat prosessin tuottoa.

Näin päädyttiin siihen, että ei pitäydytä yhdessä (muiden ehkä oikeaksi havaitsemassa) analyysissa, vaan kokeillaan useita eri lähestymistapoja järjestelmän sisäisten riippuvuuksien löytämiseksi. Näillä saaduista tuloksista pystyttäisiin sitten yhteenvedonomaaisesti muodostamaan käsityksemme bioreaktorin dynamiikasta.

Lähestymistavoiksi valittiin:

- Ajojen silmämääräinen tarkastelu
- Korrelaatioanalyysi
- Lineaarinen regressio
- Neuroverkot

Tarkastelut näillä menetelmillä suoritettiin käytännössä toisistaan riippumatta. Analyysin edetessä pyrittiin pitämään yhteisiä palavereja, jotta tutkimusta voitaisiin painottaa johonkin tiettyyn suuntaan alustavien tulosten mukaan. Tämä painotus osoittautui kuitenkin mahdottomaksi, sillä vielä noin kolme viikkoa ennen projektin määräaikaakaan minkäänlaisia merkitseviä tuloksia ei ollut saatu. Koska joka suunnalla oli vielä jotain mielenkiintoista tutkittavaa, päätettiin jatkaa loppuun asti valittujen menetelmien kanssa.

Tämän jälkeen, noin viikko ennen aikarajaa, suoritettiin yhteenveto, jonka jälkeen oli aikaa vielä nopealle lisätutkimukselle saatujen tulosten validoimiseksi ja "mallin" hahmottelemiseksi.

4.2. Ajojen silmämääräinen tarkastelu

² "Mallintaminen" on itse asiassa terminä hieman harhaanjohtava, sillä käytännössä ehdimme työssä vain analysoida tarkasteltavaa prosessia lukuisilla eri menetelmillä, eikä varsinaista mallia, puhumattakaan sen validoinnista, pystytty rakentamaan.

4.2.1. Menettelytapa

Tuotantoajojen silmämääräisessä tarkastelussa tutkittiin pääasiassa kuvaajien avulla ajojen tuottoja ja eri muuttujien kehittymistä. Ajojen tutkiminen tällä tavalla nähtiin mielekkääksi, koska tutkimukseen käytettävien ajojen määrä ei ollut lopulta yhtä suuri, kuin mitä projektia suunniteltaessa ennakoitiin. Näin silmämääräisen tarkastelun työmäärä ei ollut mahdoton. Lisäksi eri muuttujien vaikutuksen tutkimiseksi katsottiin tarpeelliseksi karsia potentiaalisia tutkimuskohteita muutamii mielenkiintoisimpiin.

Tutkimukseen otettiin mukaan kaikki mallintamiseen käytettävät AKU-, MAX- ja XAVIER-ajot, sekä lisäksi valmiiksi sähköisestä aineistosta arvotetut ajot. Tutkittavista ajoista tulostettiin seuraavien muuttujien kuvaajat:

- Vuorokausituotto
- Glukoosipitoisuus
- Laktaattipitoisuus
- pH
- Hapen osapaine

Tarkoituksena oli tällä tavalla identifioida mielenkiintoisia ilmiöitä tarkempaa tutkimista varten. Tällaisia mielenkiintoisia asioita olivat muun muassa:

- Voimakkaat muutoskohdat vasta-aineen tuotossa
- Näennäiset riippuvuudet eri muuttujien välillä
- Vasta-aineen tuoton kannalta samankaltaiset ajot

4.2.2. Havainnot

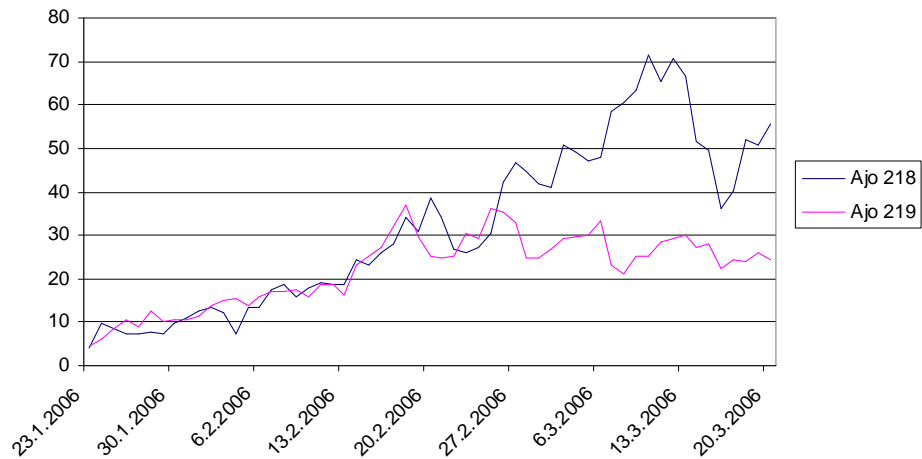
Silmämääräisen tarkastelun avulla havaittiin seuraavien ilmiöiden toistumista tarkastellussa aineistossa:

- pH:n suuri poikkeama tasapainoarvosta (noin 7.0) johtaa tuoton huononemiseen
- Laktaattipitoisuuden kasvu yli arvon 50:n mg/dl johtaa tuoton huononemiseen
- Laktaattipitoisuuden nopea muutos johtaa tuoton huononemiseen
- Glukoosi- ja vasta-ainepitoisuudessa vaikuttaa olevan jonkinasteinen negatiivinen korrelaatio

Näiden silmämääräisten havaintojen tarkempi analysointi on esitetty kappaleessa 4.3.

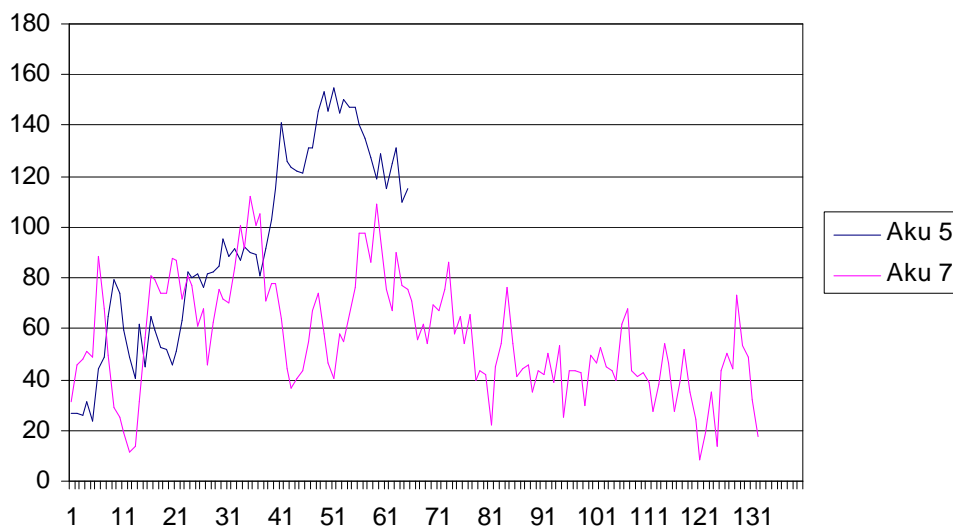
Myös vuorokausituottojen kannalta samankaltaisia ajoja löydettiin.

Vuorokausituotto ajoissa AJO218 ja AJO219 käyttäytyy aluksi hyvin samankaltaisesti, mutta poikkeaa radikaalisti noin puolivälissä ajoa. Näiden ajojen vuorokausituottoja on vertailtu allaolevassa kuvassa (Kuva 6).

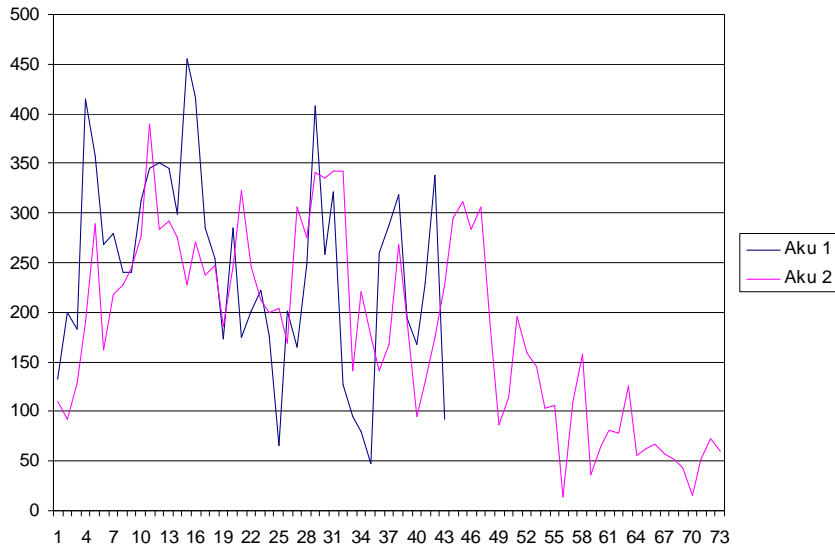


Kuva 6: Ajojen AJO218 ja AJO219 vuorokausituottojen vertailu.

Tämän ajoparin lisäksi vuorokausituottojen yhteneväisyyksiä havaittiin ajopareissa AKU5-AKU7 ja AKU1-AKU2. Näissä tuotot eivät seuraa toisiaan aivan yhtä kauniisti kuin ajojen AJO218 ja AJO219 tapauksessa, mutta niitä haluttiin kuitenkin tutkia tarkemmin. Näiden ajojen vuorokausituottojen vertailu on esitetty oheisissa kuvissa (Kuva 7 ja Kuva 8). Samankaltaisten ajojen tarkempi analyysi on esitetty kappaleessa 4.3.2.



Kuva 7: Ajojen AKU5 ja AKU7 vuorokausituottojen vertailu.



Kuva 8: Ajojen AKU1 ja AKU2 vuorokausituottojen vertailu

4.3. Tarkempi analyysi

4.3.1. Muuttujien ristikorrelaatiot

Kahden aikasarjan x_i ja y_i ristikorrelaatiofunktio $R_{XY}(i, i + j)$ on määritelty seuraavasti [3]:

$$R_{XY}(i, i + j) = E(x_i y_{i+j}). \quad (1)$$

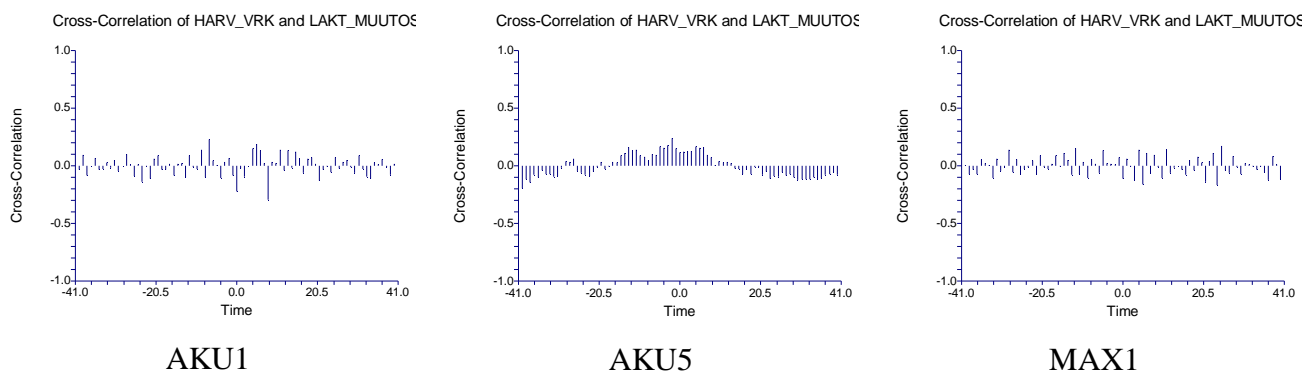
Ristikorrelaatio kertoo siis kahdessa eri muuttujassa tapahtuvien muutosten samanaikaisuudesta sekä mahdollisista viiveistä vuorovaikutuksissa. Ristikorrelaatio ei suoraan merkitse kausaliteettia, mutta sen avulla voidaan päästä jäljille tutkittavan systeemin riippuvuuksista ja osoittaa kohteita myöhemmille koeasetelmille.

Tämän suureen avulla tutkittiin aiemmin havaittuja mahdollisia riippuvuuksia. Vuorokausituoton korrelaatiota ehdittiin tutkia kahden tekijän suhteen; ensiksi tutkittiin, miten laktaattipitoisuuden muutosnopeus korreloi vuorokausituoton kanssa ja sen jälkeen, millainen korrelaatio on vuorokausituotolla ja pH-arvon poikkeamalla tasapainoarvosta 7,0. Ristikorrelaatiofunktio laskettiin viiveillä $j \in [-40, 40]$.

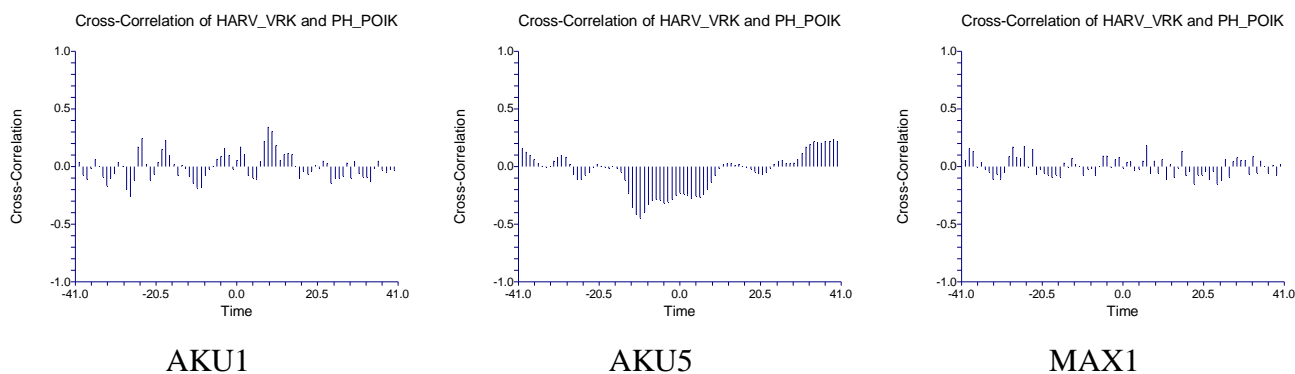
Oheisissa taulukoissa (Taulukko 2) ja (

Taulukko 3) on esitetty nämä ristikorrelaatiot kolmelle hyvälle ajolle, ajoille AKU1, AKU5 ja MAX1. Kaikkien ajojen mallinnuksessa käytettyjen ajojen vastaavat ristikorrelaatiofunktio löytyvät liitteistä 1 ja 2.

Taulukko 2: Laktaattipitoisuuden muutoksen ja vuorokausituoton ristikorrelaatiot tuotantoajoille AKU1, AKU5 ja MAX1.



Taulukko 3: pH-arvon poikkeaman ja vuorokausituoton ristikorrelaatiot tuotantoajoille AKU1, AKU5 ja MAX1.



Näiden kuvien perusteella näyttäisi siltä, että kovin suuria korrelaatioita järjestelmässä ei olisi. Ajossa AKU5 laktaattipitoisuuden muutoksella on pieni positiivinen korrelaatio vuorokausituoton kanssa, mutta ajoissa AKU1 ja MAX1 tämä korrelaatio on lähes satunnaista kohinaa. Liitteestä 1 käy ilmi, että ajo AKU5 on lähes ainoa poikkeus ajojen joukossa: muissa korrelaatio on hyvin heikkoa.

pH-arvon poikkeaman suhteen ajo AKU5 näyttää lupaavalta, mutta jälleen AKU1 ja MAX1 eivät ole kovin informatiivisia. Liitteen 2 tarkasteleminen antaa kuitenkin lisää todistetta sen puolesta, että voimakkaat pH-arvon poikkeamat ilmestyisivät samaan aikaan tuoton huononemisen kanssa; liitteen kuvista melko monessa toistuu ajon AKU5 kaltainen muoto. Korrelaatioissa oleva kausittainen rakenne viittäisi myös viiveisiin, mutta ajojen pituuksien johdosta kovin tarkkoja johtopäätöksiä ei tästä voida tehdä.

Mitään säännönmukaisuutta ajon hyvyyden suhteen ei ristikorrelaatioista ole nähtävissä. Toisaalta aika monessa ajossa molempien ristikorrelaatioiden säännönmukaisuus on samanluonteista, eli heikko ristikorrelaatio näyttää toistuvan sekä laktaatille että pH:lle. Tämä ilmiö nähdään esimerkiksi ajossa MAX5, jossa ristikorrelaatiot ovat todella minimaalisia.

4.3.2. Samankaltaisten ajojen poikkeavuudet

Aluksi tutkittiin ajoja AJO218 ja AJO219, joista lähdettiin etsimään tekijöitä, jotka ovat mahdollisesti saaneet aikaan vasta-aineen tuotannossa sellaisen notkahduksen, joka on samalla myös jäänyt pysyväksi eikä ole palannut entiselle kasvu-uralleen. Ensiksi testattiin NCSS:llä mitkä olisivat sellaisia muuttujia, jotka pystyvät selittämään tuottokäyrää mahdollisimman hyvin. Katseen kohdistaminen näihin muuttujiin auttaisi siten huomaamaan, mihin muuttujiin ja niiden arvoihin on syytä kiinnittää huomiota tuotantoajon aikana. Seuraavassa on esitetty NCSS-tuloste tästä apuestimoinnista.

Correlation Matrix Section

	DO_mg_l	Lakt_mg_dl	NH4_mmol_l	pH	Va_tuot_mg_vrk
DO_mg_l	1.000000	0.300909	0.050955	0.223679	-0.327312
Lakt_mg_dl	0.300909	1.000000	0.402001	0.846062	-0.397066
NH4_mmol_l	0.050955	0.402001	1.000000	0.595003	-0.831452
pH	0.223679	0.846062	0.595003	1.000000	-0.608674
Va_tuot_mg_vrk	-0.327312	-0.397066	-0.831452	-0.608674	1.000000

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	1220.385	381.8383	3.1961	0.002370	Reject Ho	0.880314
DO_mg_l	-21.96603	4.874199	-4.5066	0.000038	Reject Ho	0.993091
Lakt_mg_dl	0.6388925	0.2865905	2.2293	0.030140	Reject Ho	0.590204
NH4_mmol_l	-49.95059	5.541585	-9.0138	0.000000	Reject Ho	1.000000
pH	-141.1036	56.10555	-2.5150	0.015035	Reject Ho	0.694395
R-Squared	0.797717					

Correlation Matrix Section

	DO_mg_l	NH4_mmol_l	pH	Va_tuot_mg_vrk
DO_mg_l	1.000000	0.050955	0.223679	-0.327312
NH4_mmol_l	0.050955	1.000000	0.595003	-0.831452
pH	0.223679	0.595003	1.000000	-0.608674
Va_tuot_mg_vrk	-0.327312	-0.831452	-0.608674	1.000000

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	519.1926	224.4493	2.3132	0.024624	Reject Ho	0.622172
DO_mg_l	-19.82828	4.954679	-4.0019	0.000196	Reject Ho	0.975492
NH4_mmol_l	-52.67448	5.603974	-9.3995	0.000000	Reject Ho	1.000000
pH	-39.45425	33.89356	-1.1641	0.249612	Accept Ho	0.207944
R-Squared	0.778384					

Correlation Matrix Section

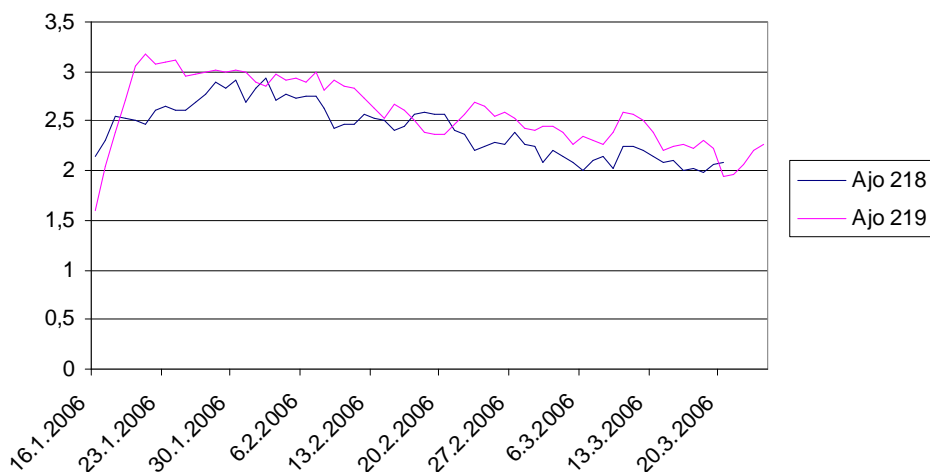
	DO_mg_l	NH4_mmol_l	Va_tuot_mg_vrk
DO_mg_l	1.000000	0.050955	-0.327312
NH4_mmol_l	0.050955	1.000000	-0.831452
Va_tuot_mg_vrk	-0.327312	-0.831452	1.000000

Regression Equation Section

Independent Variable	Regression Coefficient	Standard Error	T-Value (Ho: B=0)	Prob Level	Decision (5%)	Power (5%)
Intercept	259.2673	22.84345	11.3497	0.000000	Reject Ho	1.000000
DO_mg_l	-21.21765	4.824553	-4.3978	0.000052	Reject Ho	0.990824
NH4_mmol_l	-56.58561	4.499765	-12.5752	0.000000	Reject Ho	1.000000
R-Squared	0.772718					

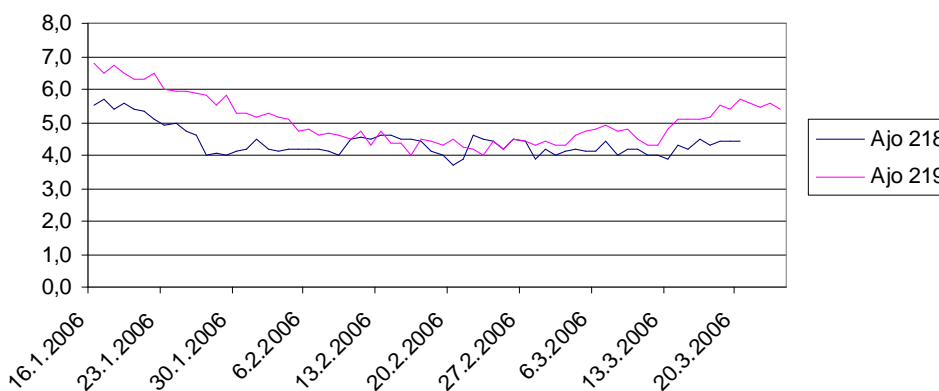
Vasta-aineen tuoton selittäjiä haettaessa jäi konkreettisesti käteen viisi muuttujaa seitsemästä, jotka mahdollisesti keskenään pystyisivät selittämään mahdollisimman hyvin vasta-aineen tuottoa. Nämä muuttujat ovat hapen osapaine, laktaatti, glutamiini, ammoniumioni ja pH. Glutamiinin määrällä ei näyttänyt olevan suurtakaan merkitystä vasta-aineen tuoton selittämisessä, vaikkakin solut tarvitsevat tätä aminohappoa jakaantuakseen. Myöskään vasta-aineen pitoisuudella ei voida kovin hyvin selittää vuorokausituottoa, vaikka vasta-ainepitoisuus onkin yksi vuorokausituoton tekijöistä.

Näitä viittä muuttujaa tarkasteltaessa lähemmin on selkeästi havaittavissa, että ammoniumionilla on vahva tilastollinen merkitsevyys selitettäessä vasta-aineen tuottoa. Ammoniumionia muodostuu aineenvaihdunnassa proteiinien kataboliassa eli hajottavassa aineenvaihdunnassa, jolle on ominaista energiaa sisältävien yhdisteiden pilkkoutuminen. Ammoniumionipitoisuuden laskiessa tuotto paranee, koska sen korkea pitoisuus häiritsee aineenvaihduntaa. Alla olevasta kuvasta näkyy kuinka paremman tuoton muodostaneen ajon AJO218 ammoniumpitoisuus on lähes koko ajon ajan pienempi kuin ajon AJO219.



Kuva 9: Ammoniumionipitoisuus ajoissa AJO218 ja AJO219.

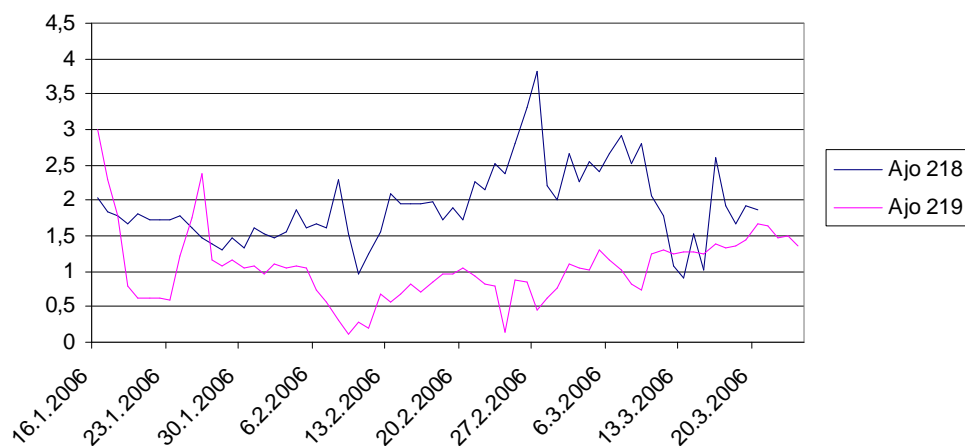
Hapen osapaineella näyttäisi olevan sellainen merkitys, että sen arvo tulisi stabiloitua likimain neljän korville, jotta se ei alkaisi vaikuttaa tuottoon negatiivisesti. Seuraavasta kuvasta on nähtävissä kuinka huonomman ajon tuotto pyrkii pienenemään mutta ajon puolivälissä, jolloin myös tuotto lähtee laskemaan, hapen osapaine lähtee yllättäen trendin lailla nousemaan.



Kuva 10: Hapen osapaine ajoissa AJO218 ja AJO219.

Toisaalta kuitenkin glutamiinin pitoisuus vaikuttaa vasta-aineen tuottoon siten, että proteiinit tarvitsevat

välttämättä tätä aminohappoa kasvaakseen. Seuraavasta kuvasta on nähtävissä myös kuinka näiden kahden ajon glutamiinipitoisuudet eroavat toisistaan. Huonommassa ajossa glutaamiinipitoisuus pienenee ajon keskivaiheen tietämillä, jolloin myös tuoton kehittyminen taantuu. Glutamiinipitoisuuden laskiessa jonkin tietyn tason alapuolelle (tässä tapauksessa n. alle 1) olisi tehtävä jotain.



Kuva 11: Glutamiinipitoisuus ajoissa AJO218 ja AJO219.

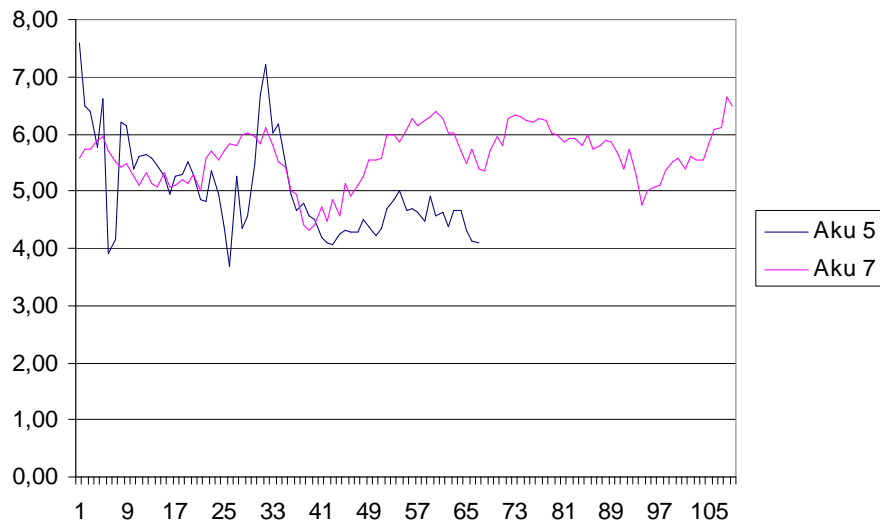
Se, miten näihin seikkoihin voitaisiin vaikuttaa pumppujen kiertonopeutta säätämällä, on kuitenkin monimutkaisempi asia. Harvestointi- ja sirkulaatiopumppujen nopeudet ovat samat 25 päivän jälkeen harvestointipumppujen käynnistämisestä. Kyseisellä ajanhetkellä ajossa AJO219 on nähtävissä asia, joka on mahdollisesti vaikuttanut tuoton pysyvään heikkenemiseen: pumppujen kiertonopeuksia on alettu säätää kiihtyvään tahtiin verrattuna ajoon AJO218, ja murroskohdassa, jossa ajon AJO219 tuotto jää selkeästi jälkeen, sen pumput käyvät n. 10 % tehokkaammin.

Ajot AKU5 ja AKU7 mukailevat toisiaan suhteellisen mukavasti aina havaintopisteeseen 35 saakka, jolloin AKU7 tuotos putoaa jyrkästi ja jää pysyvästi alhaiselle tasolle verrattuna AKU5:een. Toisaalta AKU5 joudutaan lopettamaan suhteellisen aikaisessa vaiheessa verrattuna AKU7:aan kontaminaation vuoksi.

Merkillepantavaa ajojen välillä on se, että ajossa AKU5 harvestointipumppua säädettiin alusta lähtien aktiivisesti. AKU7:ssa harvestointipumppua säädettiin vain alussa noin kerran viikossa, jonka jälkeen pumppua säädettiin pari kertaa parin viikon välein ja tuotantoajon puolivälin jälkeen pumpun nopeuteen ei tehty muutoksia.

Lisäksi AKU7:n mediapumppua ei ole säädetty tasaisen lineaarisesti suuremmalle teholle kuten ajossa AKU5 on tehty. AKU7:ssa on jäänyt kokonaan lisäämättä sirkulaatiopumpun kierroksia. Tämä on voinut osaltaan aiheuttaa huonomman tuotoksen eräajosta.

Ajosten AKU5 ja AKU7 hapen osapaineet on esitetty oheisessa kuvassa (Kuva 12).



Kuva 12: Hapen osapaineet ajoissa AKU5 ja AKU7

Hapen osapaineet mukailevat toisiaan aina näytteenottohetkeen 40 saakka, jolloin AKU7:n DO-arvo nousee voimakkaasti verrattuna AKU5:een. Samalla hetkellä myös AKU7:n tuotto heikkenee verrattuna AKU5:n tuottoon.

Kuten aiemmasta kuvasta (Kuva 8) nähdään, ajot AKU1 ja AKU2 mukailevat suurin piirtein toisiaan, mutta niiden keskihajonta poikkeaa selvästi toisistaan, kuten alla olevasta tulosteesta käy ilmi. AKU1:n keskihajonta on suurempi kuin AKU2:lla. Syynä tähän näyttäisi olevan harvestointi-, media- ja sirkulaatiopumppujen kiertonopeuksien säädöissä. AKU2:n pumppuja on säädetty kauttaaltaan rauhallisemmin verrattuna AKU1:n pumppuihin.

Summary Section of aku_1

Count	Mean	Standard Deviation	Standard Error	Minimum	Maximum	Range
43	244.5283	99.93452	15.23987	47.17887	456	408.8211

Summary Section of aku_2

Count	Mean	Standard Deviation	Standard Error	Minimum	Maximum	Range
73	182.35	95.5275	11.18065	14.16995	390.2325	376.0626

4.4. Yleinen lineaarinen malli

4.4.1. Johdanto

Yleinen lineaarinen malli on lineaarinen regressiomalli, jossa tavoitteena on selittää selitettävän muuttujan y havaittujen arvojen vaihtelu selittävien muuttujien x_1, x_2, \dots, x_p havaittujen arvojen vaihtelun avulla. Usean muuttujan lineaarinen regressio estimoi β : t yhtälöön

$$y_j = \beta_0 + \beta_1 x_{j1} + \beta_2 x_{j2} + \dots + \beta_p x_{jp} + \varepsilon_j,$$

missä β : t ovat alkuperäisiä, tuntemattomia regressiokertoimia, joiden estimaatteja kuvataan b :llä. ε_j : t

ovat virhetermejä havainnolle j . Estimoitu regressioyhtälö on seuraavanlainen:

$$\hat{y}_j = b_0 + b_1 x_{j1} + b_2 x_{j2} + \dots + b_p x_{jp}.$$

Vakio b_0 on piste, jossa regressiotaso leikkaa y-akselin ja b_i :t ovat kulmakertoimia regressiotasolla x_i :n suuntaan liikuttaessa. Nämä kertoimet kuvaavat i :n muuttujan kokonaisvaikutusta selitettävän muuttujan arvoon, kun loput selittäjät pysyvät vakiona. Estimoidun regressiomallin hyvyttä mitataan selitysasteella R^2 . Suuri osa regressioanalyysistä koostuu kuitenkin residuaalien e_j analyysistä, missä $e_j = y_j - \hat{y}_j$. Regressio-ongelma voidaan ratkaista monilla eri tekniikoilla; käytetyin on pienimmän neliösumman menetelmä (PNS). Siinä b :t valitaan siten, että residuaalien neliösumma minimoituu.

Regressiomalli on kohtuullinen ensimmäisen kertaluokan approksimaatio todellisesta mallista ja sitä voisi verrata Taylorin sarjakehitelmään. Viittaus Taylorin sarjakehitelmään kuitenkin useimmiten jättää huomiotta ”paikallisessa ympäristössä voimassaolon” -oletuksen. Toinen regressiomallissa tehty oletus on, että riippuvuus y :n ja x :n välillä on lineaarinen. Todellisuudessa ei kuitenkaan automaattisesti uskota, että tämä riippuvuus olisi suoralla viivalla kuvattavissa, mutta se on kohtuullinen ensimmäinen approksimaatio. Jos data on esimerkiksi ympyrän muodossa, regressioanalyysi ei havaitse riippuvuutta. Tästä syystä on aina suositeltavaa piirtää jokainen selittäjä selitettävän suhteen ja katsoa, jos löytyisi poikkeamaa suoralta viivalta, poikkeavia havaintoja, vaihtelun suuruuden vaihteluja tai muita poikkeuksia, joita voi ilmetä.

4.4.2. Regressiomallin standardioletukset

Mallista tehdään tavallisesti kuusi oletusta, joita kutsutaan yleistä lineaarista mallia koskeviksi standardioletuksiksi, joiden voimassaolo takaa sen, että ns. tavanomaisia estimointi- ja testausmenetelmiä saa käyttää mallin analysointiin.

- 1) Selittäjien x_i arvot x_{ji} ovat kiinteitä eli ei-satunnaisia vakioita, $j = 1, 2, \dots, K, n, i = 1, 2, \dots, K, p$
- 2) Selittäjien välillä ei ole lineaarisia riippuvuuksia.
- 3) $E(\varepsilon_j) = 0, j = 1, 2, \dots, K, n$
- 4) $Var(\varepsilon_j) = \sigma^2, j = 1, 2, \dots, K, n$
- 5) $Cor(\varepsilon_j, \varepsilon_l) = 0, j \neq l$
- 6) $\varepsilon_j \sim N(0, \sigma^2), j = 1, 2, \dots, K, n$

4.4.3. Käyräviivaisten riippuvuuksien esittäminen

Selitettävän ja selittävien muuttujien välinen käyräviivainen riippuvuus voidaan mallintaa luomalla uusia selittäjiä malliin alkuperäisistä selittäjistä neliöimällä ja joskus kuutioimalla. Esimerkiksi regressiomalli

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

voidaan laajentaa malliksi

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 \\ &= \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \beta_5 Z_5 \end{aligned}$$

Huomattakoon, että tämä malli on edelleen additiivinen uusien selittäjien suhteen.

4.4.4. Aikasarja-aineistojen käyttöön liittyviä ongelmia

Aikasarja-aineistojen käytössä on useita ongelmia. Ensinnäkin aikasarjatutkimusten tulokset ovat herkkiä valitulle tutkimusajankohdalle (aikariippuvuus). Muutaman havainnon lisäys tai vähennys aineistossa saattaa muuttaa tuloksia merkittävästi. Lisäksi aikasarja-aineistot ovat usein lyhyitä vuosiaineistoja (tässä tapauksessa muutaman kuukauden mittaisia), jolloin havaintoja on liian vähän, jotta saataisiin estimoitua tarkasti parametrien arvot keskivirheineen. Samoin aikasarjatutkimukset ovat herkkiä selittävien muuttujien, funktiomuodon ja tilastollisten menetelmien valinnalle.

Muita ongelmia aikasarja-aineistotutkimuksissa ovat virhetermien autokorrelaatio ja heteroskedastisuus sekä selittävien muuttujien voimakas multikollinearisuus. Autokorrelaatio (myös sarjakorrelaatio) tarkoittaa, että virhetermien välinen korrelaatio ei ole nolla, mikä on usein oletuksena eri regressiomenetelmissä. Vaikka esimerkiksi pienimmän neliösumman menetelmä (PNS) tuottaa harhattomia estimaattoreita autokorrelaation esiintyessä, se on tehoton menetelmä, koska on olemassa menetelmiä, jotka tuottavat myös harhattomia tuloksia, mutta pienemmällä keskivirheillä ja ovat siten tehokkaampia. F- ja t-testit eivät enää päde mallin ja kertoimien hypoteeseille eivätkä kertoimien estimaattien luottamusväleille. Positiivisella autokorrelaatiolla jäännöstermin hajonnan estimaatti (MSE) saattaa olla vakavasti aliarvioitu. Tästä seuraa kertoimien keskivirheiden aliestimointi, osittaisten t-testien arvojen pienenemistä (osoittaa valheellisesti kertoimen tilastollista merkitsevyyttä) ja kertoimien estimaattien luottamusvälien liiallista lyhentymistä. [4]

Myös selittävien muuttujien välillä saattaa olla voimakasta riippuvuutta esimerkiksi aikatrendin vuoksi. Mikäli selittävät muuttujat ovat liiaksi korreloituneita keskenään, multikollinearisuusongelmaa saattaa esiintyä tutkimuksessa. Greene [6] listaa joukon ongelmia, joita saattaa esiintyä empiirisissä tuloksissa, mikäli selittävät muuttujat ovat voimakkaasti korreloituneita:

- 1) pienet muutokset aineistossa saattavat aiheuttaa suuria vaihteluita estimoitujen parametrien arvoissa
- 2) vaikka selitysaste R^2 olisi suuri, yksittäiset kertoimet eivät ole merkitseviä suurista keskivirheistä johtuen
- 3) kertoimet saattavat olla vääränmerkkisiä ja epätodennäköisen suuria.
- 4) kertoimien t-testiarvot saattavat pienentyä ja aiheuttaa vääriä merkitsemättömiä p-arvoja. [4],[5]

Heteroskedastisuusongelma esiintyy, kun virhetermien varianssi ei säily vakiona koko tarkasteluperiodia ja silloin on mahdollista käyttää muita menetelmiä kuin PNS, jotta harhattomien estimaattien keskivirheet saataisiin pienemmiksi.

4.4.5. Regressiomallin estimointi datasta

Regressiomallin estimoinnin lähtökohdat eivät ole työssä erityisen hyvät. Datankeräystaajuus on ollut matala, jos verrataan analyysiin käytettävien selittäjien määrää olemassa olevien havaintojen määrään ja keräyksessä on ollut kohtuullisen suuri määrä reikiä (n. 25-30 %), joita on pitänyt täyttää esikäsittelynä (tässä osiossa käytettiin lineaarista interpolointia normaalijakautuneella satunnaisvaihtelulla häirittyinä). Yleisenä lähtökohhtana regression onnistumiselle on, että jokaista selittäjää ja selitettävää muuttujaa kohden olisi ainakin viisi havaintoa käytettävissä. Suosituksena on kuitenkin, että havaintoja pitäisi olla ainakin 15-20 käytettyjä muuttujia kohden.

Toinen suurempi ongelma mallin estimoinnissa liittyy siihen, että ryhmämme tuntemus työn kohteena olevasta ilmiöstä on vähäistä. Tämä vaikeuttaa mallin rakenneosan spesifiointia oikein sekä mallin alistamista asialoogisiin tarkistuksiin. Mallin parametrit eivät välttämättä ole tulkittavissa, tai ei ole täsmällistä tietoa, ovatko parametrit oikeanmerkkisiä tai -kokoisia. Sen sijaan mallin kykyä kuvata todellisuutta voidaan kuvata datan avulla. Mallia ei pitäisi koskaan valita pelkästään tilastollisin kriteerein, vaan se pitää aina alistaa asialoogisiin tarkistuksiin.

Oikeiden selittäjien löytäminen regressiomalliin on vaikeaa, koska hyvän regressiomallin jäännöseliösumma on pieni (selitysaste on korkea), mutta minkä tahansa selittäjän lisääminen malliin pienentää (ei ainakaan kasvata) jäännöseliösummaa (kasvattaa selitysastetta). Lisäksi hyvän regressiomallin kaikki selittäjät ovat tilastollisesti merkitseviä, mutta minkä tahansa selittäjän poistaminen mallista tai lisääminen malliin saattaa muuttaa malliin jäävien tai siellä jo olevien selittäjien tilastollista merkitsevyyttä.

Kun regressiomallin rakenneosalle etsitään oikeaa spesifikaatiota, keskeisenä ongelmana on löytää malliin oikeat selittäjät. Virheet regressiomallin rakenneosan spesifioinnissa saattavat johtaa karkeisiin virheellisiin johtopäätöksiin selitettävän muuttujan ja selittäjien välisestä riippuvuudesta. Jos regressiomallista puuttuu siihen kuuluvia selittäjiä, mallin regressiokertoimien PNS-estimaattorit ovat (yleensä) harhaisia. Sen sijaan, jos regressiomallissa on turhia selittäjiä, mallin regressiokertoimien PNS-estimaattorit ovat (yleensä) tehottomia, mikä merkitsee sitä, että kertoimien varianssit ovat tarpeettoman suuria.

4.4.6. Mallinrakennus regressiokertoimien estimaattien vertailemiseksi

Mallin rakentamisessa käytettiin mallinvalintatestejä, joissa tavoitteena oli ottaa malliin mukaan kaikki tilastollisesti merkitsevät selittäjät ja sulkea mallin ulkopuolelle kaikki tilastollisesti ei-merkitsevät selittäjät. Tilastollisesti merkitsevien selittäjien lisääminen malliin ja ei-merkitsevien selittäjien poistaminen mallista mallinvalintatestien perusteella ei kuitenkaan ole ongelmatonta. Testien suoritusjärjestys saattaa vaikuttaa siihen, mikä malli tulee valituksi, koska selittäjän tilastolliseen merkitsevyyteen vaikuttaa yleensä se, mitä muita selittäjiä mallissa on testauskellä.

Mallinvalintamenetelminä käytettiin askellusta alaspäin sekä ”kaikki mahdolliset mallit” (kriteereinä R^2 ,

jäännöstermin hajonnan estimaatti MSE ja Mallowsin C_p).

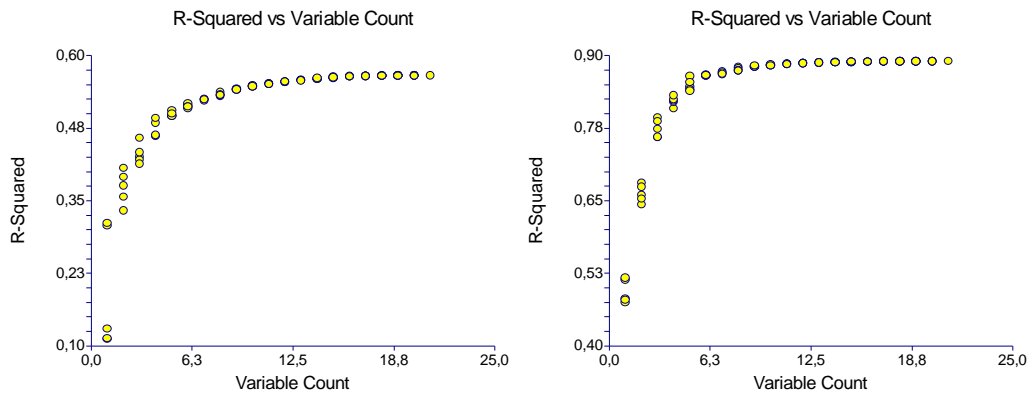
Regressiomallien kehittämisessä oli sopeuduttava erilaisiin datakokonaisuuksiin. Yksi tällainen ajokokonaisuus sisälsi täydellisen tiedon kaikista kerätyistä muuttujista, kun kahdessa muussa puuttuivat tiedot kahdesta muuttujasta (Gln ja NH_4). Viimeisessä ajokokonaisuudessa puuttuivat tiedot kolmesta muuttujasta (Gln, NH_4 ja pH). Täydellisen tiedon sisältävästä ajokokonaisuudesta yritettiin löytää malli, joka sisältäisi tärkeimmät muuttujat ja vaikutussuhteet, sekä selittäisi/ennustaisi vasta-ainepitoisuutta I_g mahdollisimman hyvin.

Mallinrakennuksessa käytettiin selitettävänä muuttujana vasta-ainepitoisuutta I_g . Sille etsittiin selittäjiä normaalien selittäjien (Hapen osapaine DO, Glukoosi, Laktaatti, Gln, NH_4 ja pH) ja näiden normaalien selittäjien ristitermien joukosta. Lisäksi selittäjien joukkoon päätettiin ottaa yksi uusi selittäjä $|pH - 7|$, koska silmämääräisissä havainnoissa oli ollut hieman tämänsuuntaisesta riippuvuudesta viitettä. Sen sijaan muut mahdolliset selittäjät (esim. kiertopumput) jätettiin pois. Selittäjäkandidaattijoukosta ei päästy malliin, joka olisi täyttänyt kaikkia regressiomallin standardioletuksia. Ongelmana oli etenkin selittäjien multikollineaarisuus, virhetermien autokorrelaatio sekä residuaalien heteroskedastisuus. Lisäksi vain muutaman kerran päästiin hyväksymään residuaalien normaalisuusoletuksen päteminen.

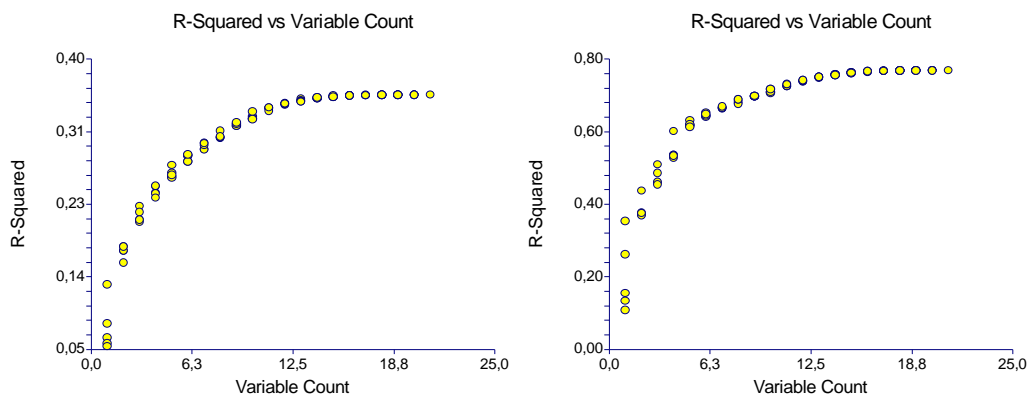
Valmiiksi sähköisenä saadusta datasta yritettiin hakea vertailukelpoisia ajoryhmiä tarkasteluun. Vertailukelpoisten ajojen määrästä johtuen tämä ei kuitenkaan täysin onnistunut, ja mallien toimivuutta päädyttiinkin lopulta tutkimaan riippumatta siitä, voitiinko ajoja vertailla keskenään, ja oliko niiden hyvyyttä ylipäänsä määritetty.

Suunnitelmana oli ottaa ainakin kaksi mahdollisimman pitkää ja normaalia ajoa tutkimuksen kohteiksi. Tavoite oli vertailemalla mallin selittäjiä sekä kertoimien merkkiä ja suuruutta löytää yhdenmukaisuuksia. Vertailuissa päädyttiin käyttämään pH-selittäjän tilalla $|pH - 7|$ -selittäjää helpomman tulkittavuuden tähden.

Tuotantoajoja vastaavien kerroinestimaattien arvot otettiin ”kaikki mahdolliset mallit” –metodilla etsien MSE:n kannalta lupaavaa selittäjäyhdistelmää ja hankkiutumalla multikollineaarisuusongelmasta eroon selittäjiä poistamalla. Tämä siksi, että autokorrelaatio oli ongelmana jokaisessa ajossa aiheuttaen t-testien pätemättömyyttä, jolloin ei ole mieltä poistaa mallista muita kuin multikollineaarisuutta aiheuttavia selittäjiä.

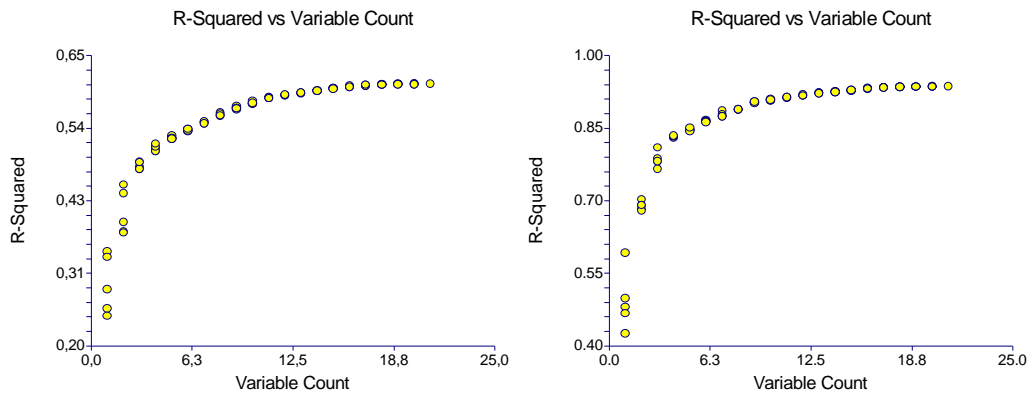


Kuva 13: Tuotantoajoista AJO212 (vasemmalla) ja AJO227 (oikealla) estimoidun mallin selitysasteen käyttäytyminen selittäjien määrän kasvaessa.

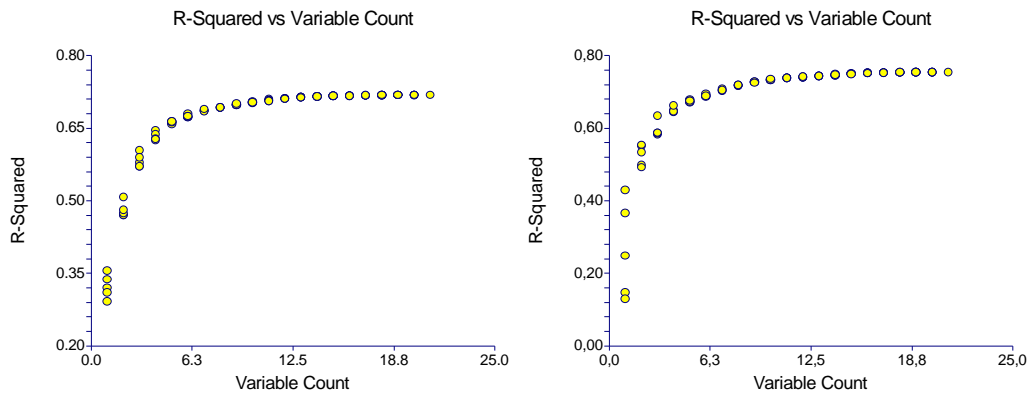


Kuva 14: Tuotantoajoista AJO222 (vasemmalla) ja KERKKO1 (oikealla) estimoidun mallin selitysasteen käyttäytyminen selittäjien määrän kasvaessa.

Oheisiin kuviin (Kuva 13 - Kuva 16) on piirretty eri tuotantoajojen selitysasteen käyttäytyminen selittäjien määrän funktiona. Mielenkiintoinen ilmiö oli, että yhdessä ajossa on jääty alle 40 % selitysasteen ja kolmessa muussakin jäädyään alle 70 %. Siis vain puolissa tutkituista ajoista oli mahdollisuus päästä yli 70 % selitysasteen, jos selittäjien määrää lisättiin kahdenkymmenen tienoille. Tähän ei kuitenkaan lähdetty, koska ei haluttu lisätä multikollineaarisuusongelmaa, eikä havaintojen määrä tuotantoajoissa olisi muutenkaan riittänyt tilastolliseen tarkasteluun niin monen selittäjän mukana ollessa. Oheisesta taulukosta (Taulukko 4) löytyy ajoista rakennettujen mallien selitysasteet ja aika huonoiksi ne ovat jääneet. Vain kaksi ajoa on saanut noin 73 % selitysasteen, yksi ajo noin 67 % selitysasteen ja loput viisi alle 60 % selitysasteen. Yleinen regressiomalli ei siis sopinut aivan odotusten mukaisesti ajoihin – edes ristitermien kanssa.



Kuva 15: Tuotantoajoista AJO218 (vasemmalla) ja AJO219 (oikealla) estimoidun mallin selityksasteen käyttäytyminen selittäjien määrän kasvaessa.



Kuva 16: Tuotantoajoista ANTERO3 (vasemmalla) ja AJO205 (oikealla) estimoidun mallin selityksasteen käyttäytyminen selittäjien määrän kasvaessa.

Standardisoidut regressiokertoimet saadaan vähentämällä jokaisesta selittäjästä sen keskiarvo ja jakamalla se keskihajonnallaan, minkä jälkeen tehdään normaali regressio. Standardisoidun regressiokertoimen yhtälö on

$$b_{j,std} = b_j \left(\frac{s_{x_i}}{s_y} \right),$$

missä s_y on selitettävän muuttujan keskihajonta ja s_{x_i} vastaa j:nnen selittäjän keskihajontaa. (Standardoimattomista) regressiokertoimista ei voi päätellä muuttujan vaikutuksen suuruutta verrattuna johonkin toiseen muuttujaan. Tämä johtuu siitä, että muuttujia mitataan eri asteikolla. Standardisoimalla muuttujat saadaan kaikki muuttujat liikkumaan samalla asteikolla välillä $[-1,1]$ ja niiden suhteellisesta selitysvuimasta voidaan tehdä johtopäätöksiä. Itseisarvoltaan suurimmat arvot vaikuttavat selitettävään voimakkaimmin.

Taulukko 4: Tuotantoajojen standardisoidut regressiokertoimet (ylempi luku) sekä tavalliset PNS-regressiokertoimet (alempi luku); alleviivattuna suurimmat ajoon vaikuttaneet selittäjäarvot ja

lihavoidut arvot eivät sisällä kertoimen luottamusvälillä nolaa.

	AJO227	AJO219	AJO205	KERKKO 1	AJO218	ANTERO3	AJO212	AJO222
Intercept	0 0,558851	0 -0,047	0 0,394673	0 -0,0476	0 -0,0228	0 1,15	0 -1,07E-01	0 0,395249
DO mg/l	0,2797 6,84E-02	0,2082 7,63E-03	-0,3608 -5,893E-02	0,5738 0,0420	0,1757 0,0106	-0,4401 -0,18836	0,5597 5,28E-02	0,2187 0,0212
Gluk mg/dl	-0,8391 -2,10E-03							-0,3212 -8,19E-04
Lakt mg/dl		0,4247 3,39E-04			0,5801 1,55E-03			
Gln mmol/l		-0,0477 -9,18E-08			0,1411 2,98E-07			
NH4 mmol/l		-0,2664 -7,21E-07						-0,2496 -1,93E-06
pH-7	0,3005 4,73E-06	-0,0472 -1,44E-07			-0,379 -1,58E-06			
DOxGluk	0,2512 4,07E-07			0,2803 2,44E-07		-0,089 -3,81E-06		
DOxLakt	0,038 1,59E-08		-0,2666 -1,655E-07				-0,2573 -6,96E-08	
DOxGln		-0,0016 -4,53E-10		-0,1099 -6,32E-08			-0,2136 -6,91E-08	0,1236 9,43E-08
DOxNH4		0,0366 6,13E-09		-0,186 -5,15E-08			0,2067 5,65E-08	
DOx pH-7		0,5276 4,48E-06		0,4767 4,53E-06				0,1584 1,38E-07
GlukxLakt			0,6266 2,661E-06	0,0587 5,99E-07		0,0958 1,06E-06	-0,1198 -9,83E-07	
GlukxGln							0,1486 1,10E-07	
GlukxNH4		0,0213 8,05E-09	-0,0981 -1,345E-07					
Glukx pH-7	0,1263 1,48E-03	0,065 2,60E-07		0,1625 9,33E-08				
LaktxGln			-0,2581 -1,441E-07		0,1251 7,96E-07			-0,3056 -5,44E-04
LaktxNH4		-0,0157 -4,48E-09	0,2024 1,450E-07	-0,2158 -1,23E-07			-0,0229 -1,37E-08	0,2247 1,72E-07
Laktx pH-7		-0,1593 -6,85E-08		-0,1372 -1,99E-06		0,0719 2,36E-06		
GlnxNH4			0,1540 3,113E-08					
Glnx pH-7				-0,3717 -1,96E-01	0,237 9,28E-02	-0,067 -1,31E-01		-0,3117 -0,44683
NH4x pH-7		-0,1124 -6,72E-07	0,1065 3,60E-02	0,1859 1,34E-06	-0,0984 -2,80E-02	0,3254 2,56E-01		
R-Squared	0,73974	0,732607	0,6694	0,5846	0,529328	0,5004	0,454853	0,31332
Ajon lopetusyy	Normaali	Ei tietoa	Epänormaali	Normaali	Normaali	Epänormaali	Normaali	Normaali
Käytettyjen havaintojen lkm	68	68	80	65	64	89	82	68
Havainnot per käytetyt muuttujat	9,71	4,85	8,88	5,41	8,00	12,71	10,25	7,56

Taulukkoon (Taulukko 4) on kerätty tuotantoajojen standardisoituja regressiokertoimia ja niistä on yritetty löytää jotain yhdenmukaisuutta selittäjien merkissä (ja voimakkuudessa). Ajot on järjestetty

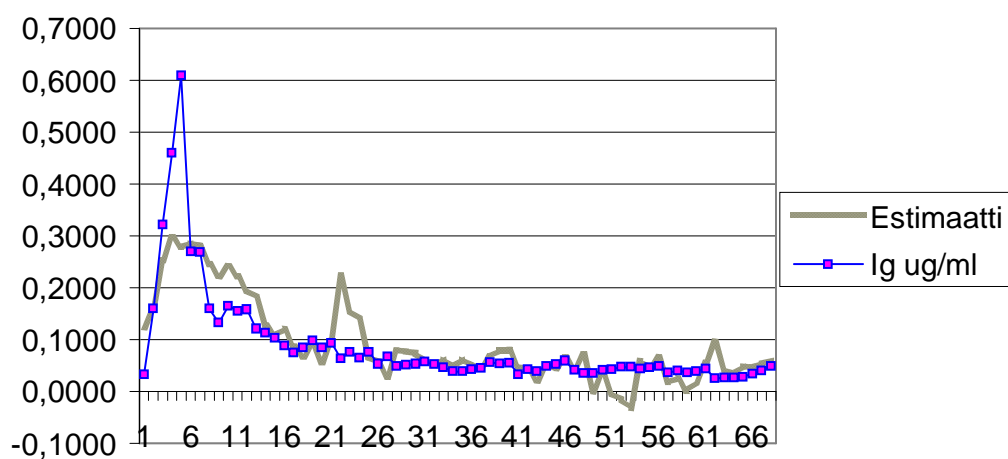
selitysasteen suuruuden mukaan. Lisäksi on merkitty ajon lopetussyy, kuinka monta havaintoa per käytetyt muuttujat (selittäjät + selitettävä) on ollut, sekä lihavoimalla merkitty kertoimet, joiden 95 % luottamusväliin ei ole kuulunut nollaa. Edellisten sekä selitysasteen suuruuden perusteella on yritetty ottaa kantaa tulosten mahdolliseen (tilastolliseen) merkitsevyyteen.

Ajoille muodostettujen mallien (MSE:n minimoiva malli, josta poistettu multikollineaarisuutta aiheuttavat selittäjät) erilaisuudella on varmasti ollut vaikutusta kerrointen suuruuteen. Jokaisessa mallissa on ollut mukana hapen osapaine (DO mg/ml). Se on saanut positiivisen kertoimen kaikissa normaalisti loppuneissa³ ajoissa. Muita eri ajoissa olleita yhdenmukaisia arvoja ovat olleet glukoosi, laktaatti ja NH₄, jotka ovat saanut samanmerkkisiä ja samaa suuruusluokkaa olevia arvoja. Lisäksi hapen osapaineen ja pH:n etäisyyden seitsemästä -yhteisvaikutus on saanut positiivisia arvoja. Erittäin varovasti tulkittuna, sillä ajot ovat saaneet hyvin erilaisia selitysasteita, näillä selittäjillä voisi olla vaikutusta vasta-ainepitoisuuteen.

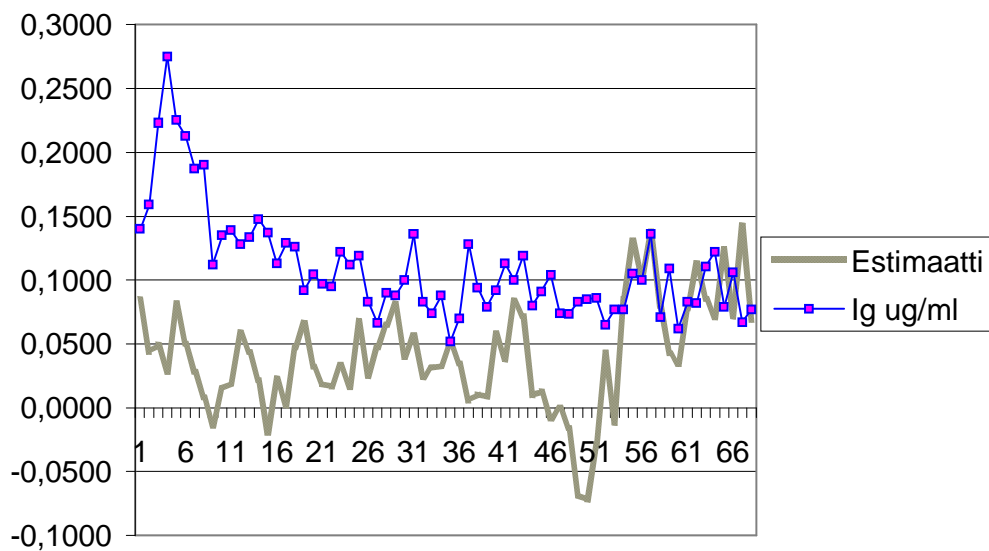
Verrattaessa parhaiten regressioon sopivia (selitysaste tyydyttävä) ajoja AJO205 ja AJO227 epäilyttävintä on se, että ns. normaali vs. epänormaali ajo ovat saaneet samoille selittäjien kertoimille eri merkit. Tässä voikin kysyä, ovatko ajot olleet niin erilaisia, vai puuttuuko mallista aitoja selittäjiä, jonka seurauksena jommankumman ajon kertoimet ovat harhaisia. Vai onko syynä ollut sittenkin ajoissa olleiden reikien täyttäminen, mikä on saattanut hävittää osan riippuvuussuhteista ajoissa (20-30 % datasta reikiä).

Tarkasteltaessa tuotantoajon AJO227 avulla estimoidun mallin sopimista ajoihin AJO227 (ks. Kuva 17), AJO212 (ks. Kuva 18) ja AJO205 (ks. Kuva 19) huomataan, että tulos on ollut heikko ajoille AJO212 ja AJO205. Niissä on ollut liian paljon heilahtelua vasta-ainepitoisuudessa verrattuna ajoon AJO227, mikä on aiheuttanut varmasti myös kertoimien erilaisuutta. Ideaalitapauksessa olisi ollut edes yksi pidempi (mielellään lähemmäs 200-300 havaintoa sisältänyt tuotantoajo), josta olisi saatu estimoitua riippuvuusmalli, jonka tilastollinen merkitys olisi vahvempi. Tätä mallia olisi sitten käytetty kaikille ajoille, minkä jälkeen olisi sitten tehty kerroinvertailu.

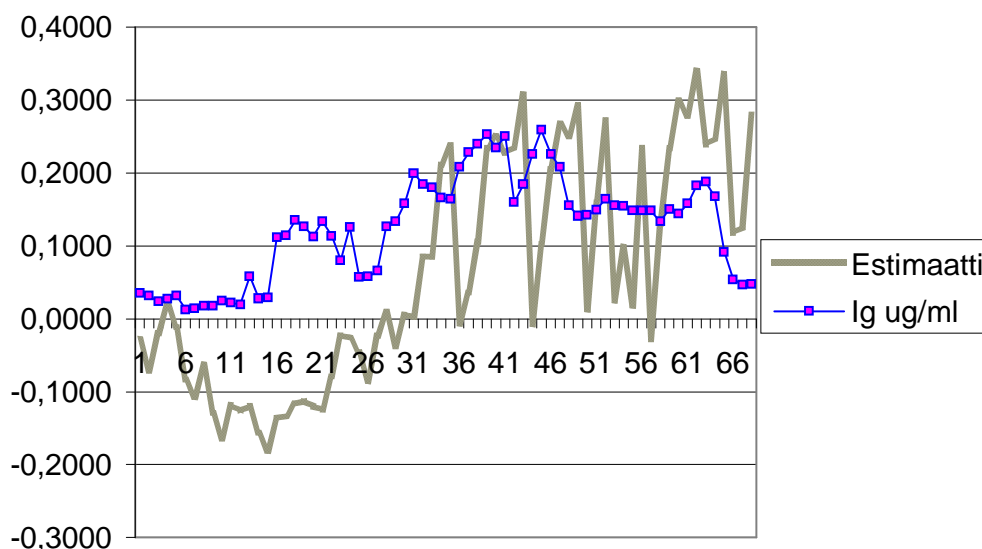
³ Ajo on loppunut normaalisti, kun se on lopetettu, kun lopputuotetta on saatu riittävästi.



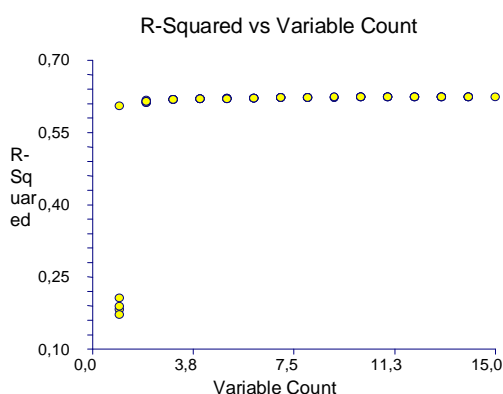
Kuva 17: Tuotantoajosta AJO227 estimoidun mallin sovitus ajoon AJO227.



Kuva 18: Tuotantoajosta AJO227 estimoidun mallin sovitus ajoon AJO212.



Kuva 19: Tuotantoajosta AJO227 estimoidun mallin sovitus ajoon AJO205.



Kuva 20: Tuotantoajosta MAX5 estimoidun mallin selitysasteen käyttäytyminen selittäjien määrän kasvaessa.

4.4.7. Ennustemallin kehittäminen eri laitteilla tehdyille tuotantoajoille

Lähes kaikista Maximizer- ja Xcell-laitteen tuotantoajoista puuttui Gln- ja Nh4-muuttujien tiedot. Lähdettiin kokeilemaan vasta-ainepitoisuus Ig:n selittämistä muuttujilla DO, laktaatti, glukoosi, $|pH - 7|$ sekä näiden ristitermit. Lisäksi otettiin mukaan ensimmäiset viipeet muuttujista DO, laktaatti, glukoosi, $|pH - 7|$. Mallista yritettiin saada vähän dynamisempi. Mallinetsinnässä käytettiin jälleen ”kaikki mahdolliset mallit” -menetelmää. MAX5-datalla (kaikkein pisin, 203 havaintoa) päästiin yksinkertaisimpaan ja selitysasteeltaan kohtuullisen hyvään malliin. Kuten oheisesta kuvasta (Kuva 20) näkyy, tarvitaan vain muutama selittäjä, jotta päästään selitysasteen maksimiin. Selittäjiksi tulivat Ig_{t-1} , DoxLakt, GlukxLakt eli tuloksena oli stokastinen differenssimalli.

Löydetyssä mallissa ei ollut yhtä pahoja ongelmia, mitä valmiiksi sähköisenä saadun datan ajoilla oli. Multikollineaarisuus ei ollut ongelma, eikä heteroskedastisuuskaan, mutta autokorrelaation suuruus oli

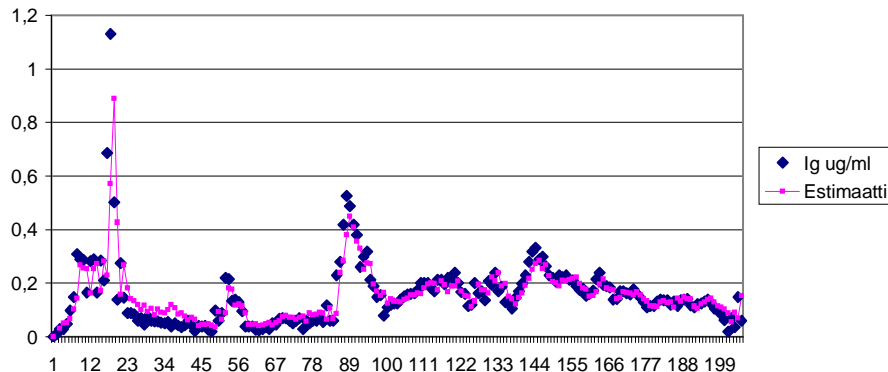
ylärajoilla. Durbin-Watson-suure oli 1,7423 ja vain toisella viipeellä oli merkitsevää korrelaatiota ($-0,201312 > \text{vertailuarvo} = 0,140372$). Stokastisen differenssimallin estimointi PNS-menetelmällä ei ole järkevää, jos residuaalit ovat autokorreloituneita, sillä kertoimien estimaatit menevät harhaisiksi. Tämän estämiseksi pitäisi käyttää estimoinnissa suurimman uskottavuuden menetelmää.

On kuitenkin toinen vaihtoehto autokorreloituneiden residuaalien ongelman korjaamiseksi nimittäin Hatanakan menetelmä. Sillä saadut estimaattorit ovat asympotoottisesti ekvivalentteja SU-estimaattoreiden kanssa, mutta eivät ole pienten otosten ominaisuuksiltaan välttämättä yhtä hyviä kuin SU-estimaattorit. Hatanakan menetelmän etuna on kuitenkin se, että siinä estimaattorit saadaan estimoimalla kolme lineaarista regressiota tavallisella PNS-menetelmällä. [7] Mallin kertoimet laskettiin sekä PNS- että Hatanakan menetelmällä, joiden tulokset ovat allaolevassa taulukossa (Taulukko 5). Kuten nähdään, saatiin samanmerkkiset ja melkolailla samansuuruiset kertoimet. Malliksi saatiin siis

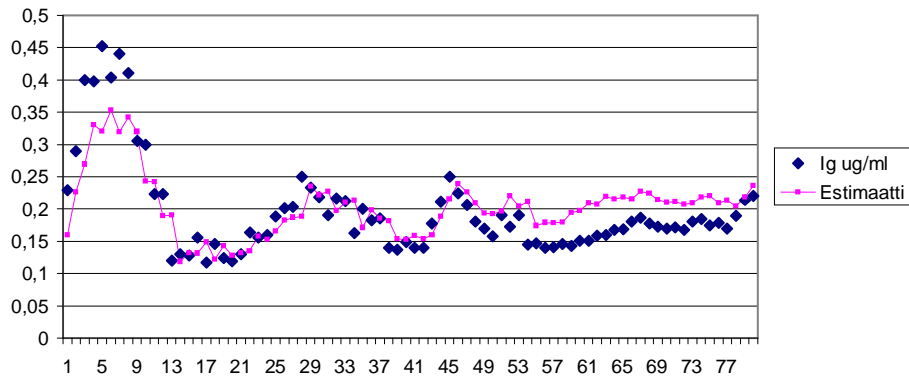
$$I_{g_t} = 0,6218973 \cdot I_{g_{t-1}} + 2,24692 \cdot 10^{-4} \cdot DO \cdot Lakt - 7,21 \cdot 10^{-7} \cdot Gluk \cdot Lakt \quad (2)$$

Taulukko 5: Maxi 6 kerrointen estimaatit sekä Hatanakan- että PNS-menetelmällä.

	Hatanaka	PNS
$I_{g_{t-1}}$	0,6218973	0,726104
Do x Lakt	2,24692E-04	2,15E-04
Gluk x Lakt	-7,21E-07	-1,73E-06

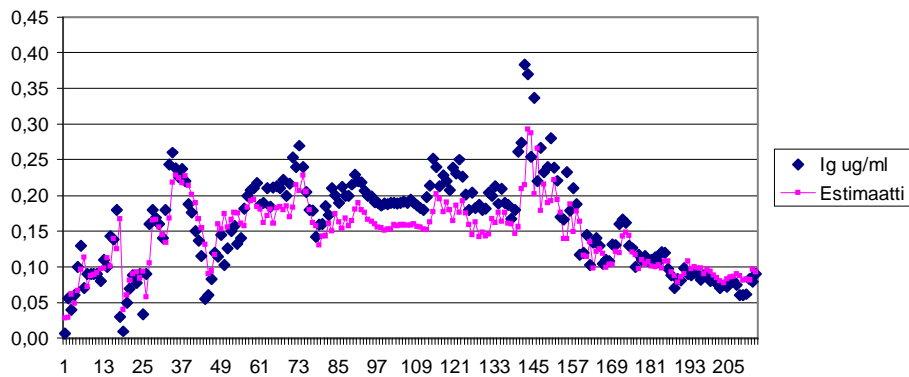


Kuva 21: Ajosta MAX5 estimoitu stokastinen differenssimalli sovitettuna alkuperäiseen dataan.

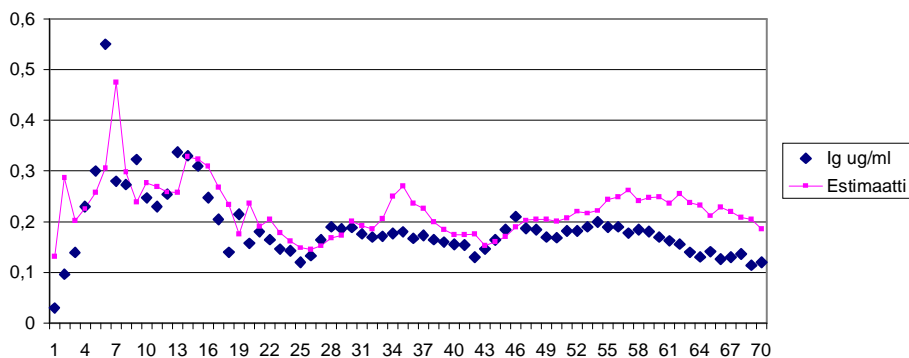


Kuva 22: Ajosta MAX5 estimoitu stokastinen differenssimalli sovitettuna MAX4-ajoon.

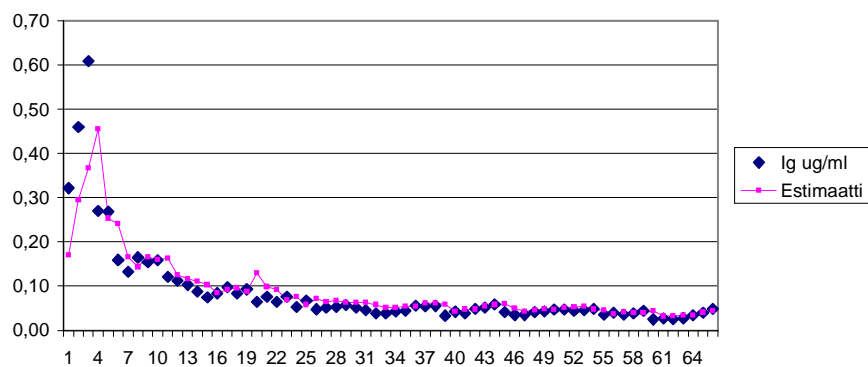
Mallin (2) ennustavuutta lähdettiin kokeilemaan ja tulokset näkyvät oheisissa kuvissa (Kuva 21 - Kuva 26): selittäjillä on pystytty kuvaamaan kohtuullisella tarkkuudella vasta-ainepitoisuutta Ig. Kaikkein huonoimmin malli näyttäisi sopineen Acu-laitteen ajoon AKU5, mutta se on sopinut huomattavasti paremmin ajoon AJO212 kuin edellisessä luvussa ajon AJO227 pohjalta luotu malli (vrt. Kuva 18 ja Kuva 26). Stokastinen differenssimalli (2) näyttäisikin olevan sopeutuvampi myös heilahtelevampaan aikasarjadataan.



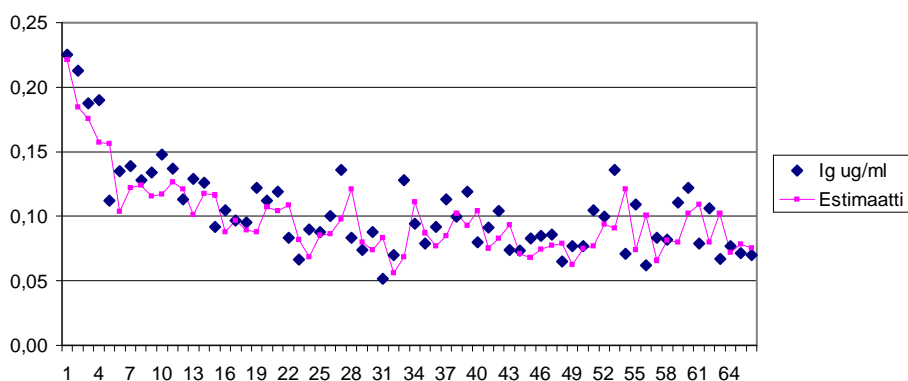
Kuva 23: Ajosta MAX5 estimoitu stokastinen differenssimalli sovitettuna ajoon XAVIER4.



Kuva 24: Ajosta MAX5 estimoitu stokastinen differenssimalli sovitettuna ajoon AKU5.

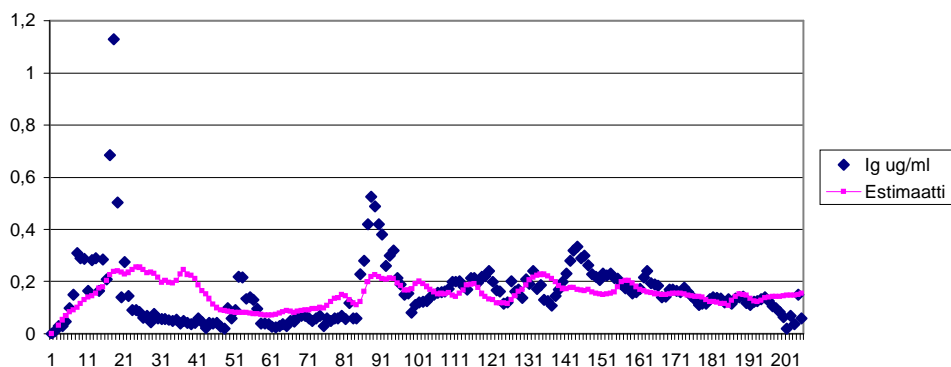


Kuva 25: Ajoin MAX5 estimoitu stokastinen differenssimalli sovitettuna ajoon AJO227.



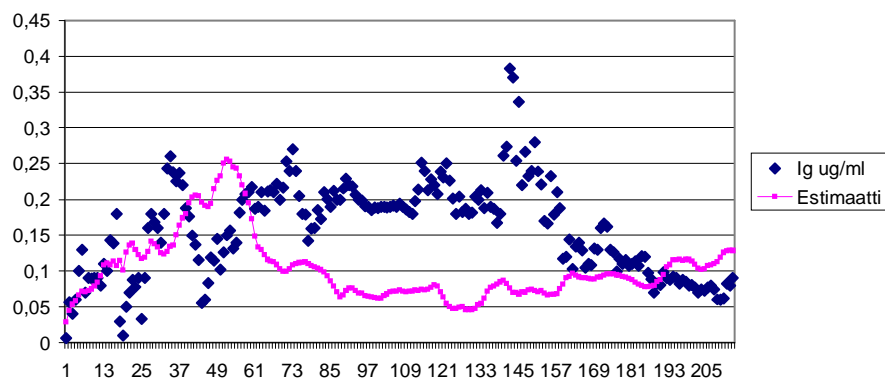
Kuva 26: Ajoin MAX5 estimoitu stokastinen differenssimalli sovitettuna ajoon AJO212.

Mallilla (2) lähdettiin etsimään myös jonkinikäistä ennustushorisonttia eli suositusta kuinka kauas sillä voisi olla mahdollista ennustaa, jos tiedettäisiin (tai osattaisiin tarkasti ennustaa) Hapen osapaineen, Glukoosin ja Laktaatin arvot tulevaisuudessa. Oheisiin kuviin (Kuva 27 ja Kuva 28) on kerätty kuvaajat mallin käyttäytymisestä tarkoilla (tiedetyillä) ajodatoilla. Ensimmäisellä ajanhetkellä on annettu Ig:lle oikea alkuarvo, mutta sen jälkeen on käytetty joka ajanhetkellä aina stokastisen differenssimallin (2) antamaa Ig:n estimaattia.



Kuva 27: Ajoin MAX5 estimoidun stokastisen differenssimallin ennustehorisontti (sovitettuna

alkuperäiseen dataan).



Kuva 28: Ajoista MAX5 estimoidun stokastisen differenssimallin ennustehorisontti (sovitettuna ajoon XAVIER4).

Malli (2) pärjää kohtuullisesti alkuperäisellä datalla, josta se on estimoitu (ks. Kuva 27). Mutta heti käytettäessä sahaavampaa Ig-aikasarjaa (ks. Kuva 28) totuus paljastuu: malli pystyy ennustamaan jollakin asteella enintään noin 10 päivän päähän, minkä jälkeen se voi antaa aivan vääriä arvoja. Tämän vuoksi voidaan päätellä, että paljon systeemiin liittyvästä tiedosta on sisällytetty stokastisessa differenssimallissa (2) Ig_{t-1} -muuttujaan. Tämän vuoksi tulokset kertomien todellisesta merkityksestä vasta-ainepitoisuudelle Ig on otettava suurella varauksella huomioon.

4.4.8. Regressioanalyysi ajojen erotuksille

Tämä osio suoritettiin käyttämällä lineaarisella interpolaatiolla täytettyjä Maximizer-laitteen ajotiedostoja. Parhaiten eri muuttujien vaikutuksia ajon onnistumiseen voisi tarkastella, jos käytettävissä olisi ideaalinen ajo kaikkine muuttujineen. Muiden ajojen muuttujien arvoja pystyttäisiin vertaamaan ideaaliajon vastaaviin ja kyettäisiin muodostamaan mallin, jossa selittäjinä olisivat muuttujien väliset erotukset. Tästä nähtäisiin suoraan millä tekijällä on suurin vaikutus ajon huonoon tuottotasoon.

Aineistona käytettiin MAX-ajoja. Ideaaliajoa ei ollut käytettävissä, joten ajojen arvotuksessa saatu parhain ajo valittiin ns. verrokkiajoksi. Myös tällä tavalla pystytään toki tarkastelemaan ajojen tuotantoeroihin vaikuttavia tekijöitä. Paras Maximizer-laitteen ajo oli MAX1. Regressioanalyysi suoritettiin MAX1- ja muiden Maximizer-laitteen ajojen aikasarjojen muuttujien erotuksille. Ensin määritettiin kunkin reaktorin vasta-aineentuotantotaso kertomalla vasta-ainepitoisuus harvestointipumpun kiertonopeudella. Regressioanalyysissä oli selitettävänä muuttujana MAX1-ajon ja kunkin muun MAX-ajon vasta-ainetuotannon tasojen erotus. Selittäjinä käytettiin hapen osapaineen (DO), glukoosi- ja laktaattipitoisuuksien sekä pH-arvojen erotuksia. Selittäjinä olivat lisäksi kaikkien edellä mainittujen muuttujien ristitermit. Jotta prosessin mahdolliset viipeet olisi saatu otettua huomioon, selittäjien määrä kaksinkertaistettiin lopuksi ottamalla mukaan kaikki termit yhdellä päivällä viivästettynä. Aineistosta poistettiin alusta ne päivät, joissa harvestointipumpun nopeus oli ollut 0.

Regressiomallin estimointi aloitettiin käyttämällä ajojen MAX1 ja MAX2 aikasarjojen erotusta. Ensin poistettiin tilastollisesti merkityksettömät selittäjät käyttämällä 15 % -merkitsevyystasoa. Selittäjien poistamisen jälkeen tarkasteltiin selittäjien multikollineaarisuutta. Selittäjien välillä oli erittäin voimakasta multikollineaarisuutta, josta pyrittiin eroon vähentämällä selittäjien määrää. Lopullisiksi selittäjiksi saatiin termit

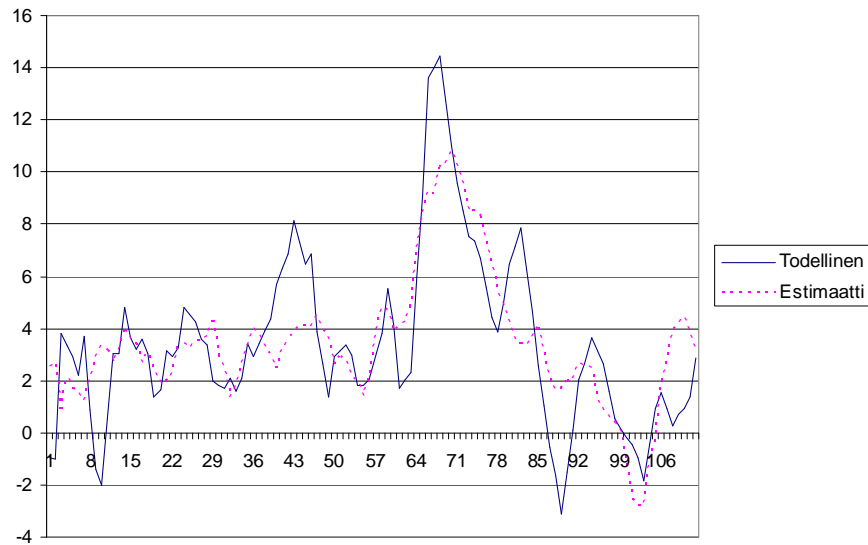
- pH
- Gluk*Lakt
- Lakt*pH
- $DO_{t-1} * Lakt_{t-1}$
- $DO_{t-1} * pH_{t-1}$
- $Gluk_{t-1} * pH_{t-1}$

Kunkin regressioanalyysin selittäjien kertoimet on esitetty oheisessa taulukossa (Taulukko 6). Niistä voidaan havaita, että pH:n kerroin on selvästi suurin. Tämä viittaisi siihen, että pH:lla olisi suuri vaikutus vasta-ainetuotannon tasoon, joten sitä tulisi seurata tarkasti. Suuret poikkeamat optimaalisesta tasosta saattavat aiheuttaa suuren vasta-ainetuotannon alenemisen. Ajojen välisistä suurista tuotantoeroista huolimatta regressiokertoimet ovat melko samanlaisia kaikille ajoille. Tulosten tulkinnassa täytyy ottaa kuitenkin huomioon, että muutamille aineistoille selittäjien multikollineaarisuus saattaa aiheuttaa virhettä kertoimissa. Lisäksi residuaalien autokorrelaatio oli melko voimakasta. Tämä johtaa siihen, että regressiokertoimien pienimmän neliösumman estimaatit eivät ole enää parhaita lineaaristen ja harhattomien estimaattien joukossa. Edellä mainituista huolimatta tulokset ovat varmasti ainakin suuntaa-antavia ja antavat perustelun seurata pH:n muutoksia tarkemmin. Myös mallia, jossa selitettävänä muuttujana oli vasta-ainepitoisuus vasta-ainetuotannon tason asemesta, testattiin, mutta tulokset olivat heikompia ja tilastollinen päättely niiden pohjalta erittäin vaikeaa.

Taulukko 6: Eri aineistojen regressiokertoimet. pH:lla näyttäisi olevan suuri vaikutus vasta-ainetuotantotasoon.

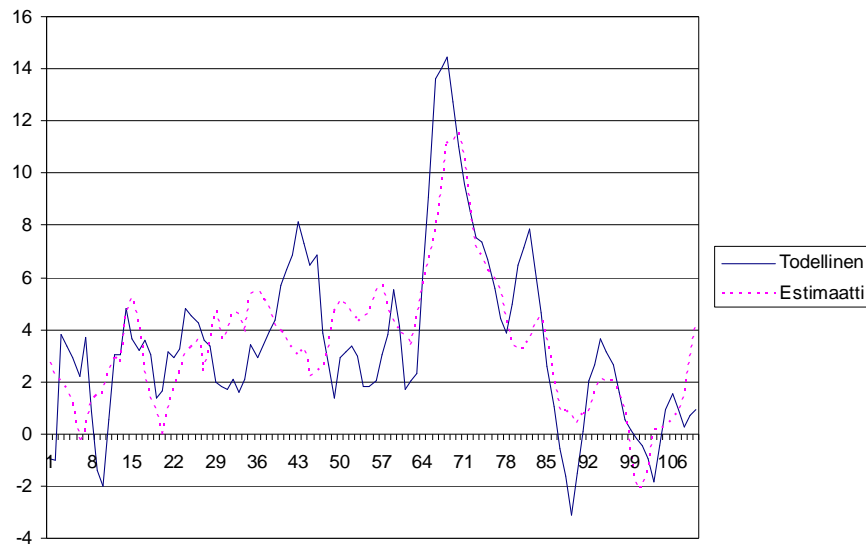
	MAX1- MAX2	MAX1- MAX3	MAX1- MAX4	MAX1- MAX5	MAX1- MAX9	MAX1- MAX10
pH	17,62	10,76	25,24	6,53	-1,88	17,36
Gluk*Lakt	8,37E-04	1,46E-03	3,86E-05	-1,35E-04	1,92E-04	6,00E-04
Lakt*pH	-3,96E-02	-5,32E-02	1,90E-03	2,96E-03	4,64E-03	-2,65E-02
$DO_{t-1} * Lakt_{t-1}$	-1,07E-02	-1,48E-02	-2,72E-02	-1,56E-02	-1,23E-02	-1,58E-03
$DO_{t-1} * pH_{t-1}$	0,18	0,11	0,31	1,39E-02	0,10	0,04

$\text{Gluk}_{t-1} * \text{pH}_{t-1}$	-1,29E-02	-1,34E-02	-3,77E-03	-1,03E-02	3,06E-03	-3,56E-03
---------------------------------------	-----------	-----------	-----------	-----------	----------	-----------



Kuva 29: Ajojen erotuksille muodostetun regressiomallin sovitus ja alkuperäinen aineisto.

Edellä olevassa analyysissä ei ole lainkaan otettu huomioon sirkulaatio- ja mediapumppujen kiertonopeuksia. Niiden ottaminen mukaan analyysiin voisi johtaa parempiin tuloksiin. Pumppujen kiertonopeudet ovat kuitenkin olennaisia prosessin ohjaamisessa. Mediapumpun kiertonopeus pyrittiin ottamaan huomioon muodostamalla glukoosin syöttöä kuvaavan muuttujan kertomalla reaktoriin syötettävän ravintoliuoksen glukoosipitoisuutta mediapumpun kiertonopeudella. Sirkulaatiopumpun kiertonopeus pyrittiin ottamaan huomioon kertomalla hapen osapaineen, glukoosi- ja laktaattipitoisuuden sirkulaatiopumpun kiertonopeudella. Näiden operaatioiden jälkeen muodostettiin lineaarinen regressiomalli samalla tapaa kuin edellä, eli laskettiin ajon MAX1- ja muiden ajojen muuttujien erotukset ja muodostettiin malli näille. Mukana olivat taas ristitermit sekä yhdellä päivällä viivästetyt termit. Tulokset olivat jälleen heikkoja. Käytettävänä olleilla selittäjillä ei pystytty muodostamaan tilastollisesti merkitsevää mallia ennako-odotusten vastaisesti. Muun muassa selitysasteet jäivät hyvin matalalle tasolle.



Kuva 30: Pumppujen kiertonopeudet huomioivan mallin sovite ja alkuperäinen aineisto.

4.4.9. Regressioanalyysi muuttujien muutoksille

Koska edellisissä malleissa yhtenä ongelmana oli residuaalien suuri autokorrelaatio, testattiin regressiomallia muuttujien muutoksille. Eli kaikkia aikasarjoja differoitiin yhden kerran ja muodostettiin lineaarinen malli, joka pyrki selittämään yhden ajon vasta-ainetuotantotason muutosta seuraavilla selittäjillä: hapen osapaineen DO muutos, glukoosi- ja laktaattipitoisuuksien muutos, pH:n muutos sekä näiden muuttujien differoimattomien arvojen ristitermien muutokset. Malleja yritettiin muodostaa muutamille ajoille (MAX1, MAX2 ja MAX10), mutta tulokset olivat huonoja. Mallien selitysasteet jäivät 10 % tienoille, joten tämä lähestymistapa hylättiin toimimattomana. Aineistona käytettiin myös pumppujen kiertonopeuksilla skaalattuja muuttujia, mutta lopputulos oli sama. Mitään tilastollisia päätelmiä ei ollut mahdollista tehdä.

4.4.10. Pohdintaa

Huolimatta erittäin useiden erilaisten lineaaristen mallien testaamisesta, kovinkaan toimivaa mallia ei saatu muodostettua. Tiesimme, että biologinen systeemi on hyvin epälineaarinen, joten malleihin lisättiin ristitermejä. Erilaisia muuttujien ristitermiyhdistelmiä on kuitenkin loputon määrä, joten tyydyimme vain ensimmäisen kertaluvun ristitermeihin ilman muuttujien muunnoksia. Muuttujien muunnokset olisivat saattaneet helpottaa hyvän mallin löytämistä, mutta niiden löytäminen olisi ollut varmasti aikamoista arpapeliä, joten siihen ei tämän projektin puitteissa lähdetty. Lisäksi bioprosessissa esiintyy varmasti viipeitä, joiden määrittäminen on hankalaa. Ongelmat regressiomallin perusoletusten täyttämässä viittaisivat väärin identifioituun mallin rakenteeseen. Tämä on seurausta vääristä tai puutteellisista selittäjistä. Tätä ongelmaa kannattaa pohtia, jos asian tutkimista jatketaan. Varmistamalla parempi tuntemus itse bioprosessista voi helpottaa sopivien selittäjien löytämistä.

4.5. Bioreaktorin mallintaminen neuroverkolla

Yhtenä osana projektityötä bioreaktorin toimintaa pyrittiin mallintamaan neuroverkolla. Koska aikasarjatieto oli kerätty useasta eri laitteesta sekä useasta eri solulinjasta, olisi käytännössä pitänyt rakentaa useampi verkko. Tämä olisi kuitenkin muodostunut liian työlääksi, joten pyrittiin rakentamaan malli suurimmasta vertailukelpoisesta datasetistä. Täksi osoittautui Maximizer-laitteesta kerätyt 11 ajon mittaukset. Verkon opettamiseen käytettiin kuutta ajoa ja sen todentamiseen kahta. Loput kolme ajoa osoittautuivat käyttökelvottomiksi.

4.5.1. Neuroverkko

Neuroverkko on laskentamenetelmä, joka perustuu ihmisaivojen tapaan prosessoida tietoa. Se koostuu toisiinsa kytketyistä yksinkertaisista laskentayksiköistä – neuroneista – jotka yhdessä muodostavat verkon ulostulon. Toisin kuin perinteisemmissä laskentamenetelmissä, neuroverkoissa tiedonkäsittely tapahtuu rinnakkaisesti. Usean neuronin yhteistyö mahdollistaa varsin monimutkaisten prosessien mallintamisen. Tällä on kuitenkin hintansa; monimutkaisten systeemien mallintaminen vaatii yleensä suuren määrän neuroneja, jolloin verkon tarkka toiminta jää hämärän peittoon. Neuroverkkojen muodostamat mallit ovatkin ns. blackbox-malleja. Tämä tuo tiettyjä vaikeuksia myös tulosten tulkintaan.

Neuroverkot ovat opetettavia systeemejä. Tämä tarkoittaa sitä, että kun verkko luodaan, se pitää opettaa mallintamaan haluttua systeemiä valmiiksi kerätyllä datalla. Tämä käsittää systeemin toimintaa ohjaavia suureita sekä ulostulosuureet. Myös online-oppiminen on mahdollista. Tätä ominaisuutta käytetään paljon mm. säätötekniikassa, jossa neuroverkkoon perustuvan säätimen toiminta paranee ajan myötä sen oppiessa prosessista koko ajan lisää.

Tavoitteena on opettaa verkolle systeemin yleisiä lainalaisuuksia, jolloin se osaa mallintaa myös tilanteita, joita ei esiinny testidatassa. Tähän vaikuttaa opetukseen käytettävän tiedon lisäksi verkon koko ja muoto. Neuronien lukumäärän ei tule olla liian suuri, sillä tällöin vaarana on, että verkko ”muistaa” testidatan (overfitting), eikä siten oikeasti kuvaa systeemin käyttäytymistä.

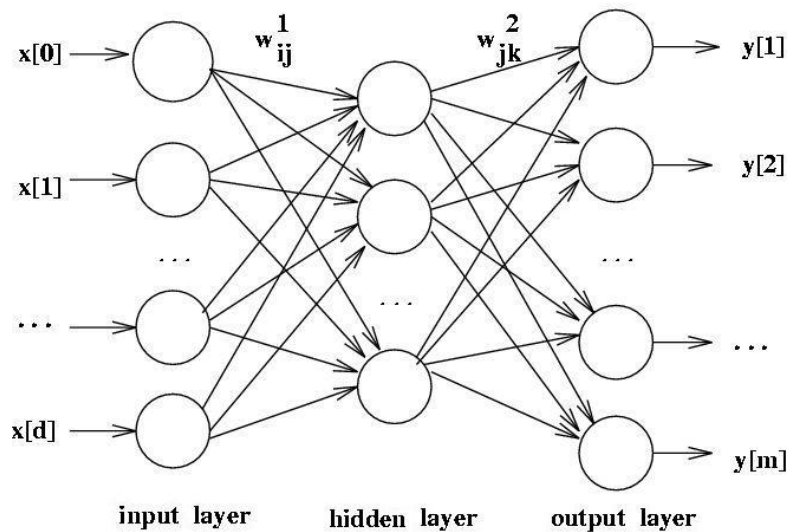
Verkon käyttäytymiseen vaikutetaan muuttamalla neuronien välisiä painoja. Erilaisia oppimisalgoritmeja on kehitetty useita, niistä kuuluisin on backpropagation-algoritmi, jota käytetään myös tässä työssä.

Erilaisia neuroverkkoarkkitehtuureja on monenlaisia. Näistä yksinkertaisin ja eniten käytetty on feedforward-verkko (Kuva 31). FF-verkko koostuu sisäänmenokerroksesta (input layer), ulostulokerroksesta (output layer) sekä yhdestä tai useammasta piilokerroksesta (hidden layer). Näistä kussakin voi olla yksi tai useampi neuroni.

Jokainen sisäänmeno ohjataan jokaiselle ensimmäisen piilokerroksen neuronille. Ensimmäisen piilokerroksen neuronit summaavat jokaisen sisäänmenon ja suorittavat siirtofunktionsa mukaisen laskutoiminnon. Tämän jälkeen ensimmäisen kerroksen neuronien ulostulo ohjataan kullekin seuraavan kerroksen neuronille ja niin edelleen. Myös ulostulokerroksen neuronit tekevät siirtofunktionsa mukaisen laskutoiminnon.

Siirtofunktio voi olla mikä tahansa mielivaltainen funktio. Usein niiden ulostulo kuitenkin rajoitetaan

välille $[-1,1]$ tai $[0,1]$. On olemassa muutamia de facto-siirtofunktioita, kuten hyperbolinen tangenti, askelfunktio, lineaarinen sekä Gaussin kellokäyrä.



Kuva 31: Feedforward-tyyppinen neuroverkko, jossa on sisään- sekä ulostulokerroksen lisäksi yksi piilokerros .

4.5.2. Verkon muodostaminen

Tärkeä vaihe neuroverkkoa suunniteltaessa on sisäänmenojen valinta. Tässä työssä tätä rajoitti kuitenkin testidatan määrä sekä laatu. Olemassa olevasta aikasarjadatasta voitiin käytännössä käyttää viittä bioreaktorin parametria (hapen osapaine, glukoosi, laktaatti, pH sekä vasta-ainepitoisuus), sillä muut olivat liian puutteellisia. Näiden lisäksi bioreaktorin toimintaan osallistuu 3 pumppua (harvestointipumppu, mediapumppu ja sirkulaatiopumppu), joiden aikasarjadata oli myös käytettävissä.

Systeemin ulostulona käytetään kunkin ajon kumulatiivista tuottoa. Kuten muukin data, myös kumulatiivinen tuotto on kirjattu vain noin viikon välein. Vastoin kuin muut mittaukset, ulostulon aikasarja täydennettiin sovittamalla siihen neljännen asteen polynomi.

Koska kumulatiivinen tuotto oli luonnollisestikin riippuvainen ajasta, lisättiin systeemin vielä yksi sisäänmeno: aika.

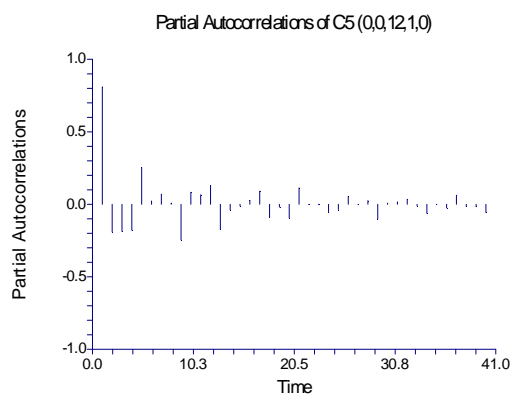
Lisäksi huomioon piti ottaa mahdolliset viiveet. Mikäli viiveitä esiintyy, pitää neuroverkon sisäänmenoiksi lisätä viivästettyjä signaaleita. Mahdolliset viiveet voidaan havaita parhaiten aikasarjojen korrelaatiofunktioista. Tätä vaikeutti kuitenkin olemassa olevien aikasarjojen lyhyys. Koska käytössä olleet kahdeksan ajoa olivat kuitenkin samasta reaktorista ja kyseessä oli sama solulinja, tutkittiin vain yhtä, eniten mittauksia sisältänyttä aikasarjaa. Vaarana tällöin kuitenkin oli, että mukaan tulisi vain tässä ajossa esiintyviä poikkeavuuksia, jolloin viiveet eivät olisi olleet yleisesti päteviä.

Malliin voidaan tosin ottaa mukaan tilastollisesti merkityksettömiäkin sisäänmenoja. Turhia sisäänmenoja

tulee kuitenkin välttää, jottei verkosta tule turhan monimutkainen. Merkitsevien viiveiden poisjääminen ei tietenkään ole toivottavaa.

Kun tutkittiin aikasarjojen auto- sekä osittaisautokorrelaatioita, autokorrelaatiofunktioista huomattiin, että aikasarjat eivät olleet stationaarisia, vaan niissä on selvä lineaarinen trendi. Trendit poistettiin differoimalla niitä kertaluvulla 1.

Kun aikasarjoista oli saatu stationaarisia, voitiin viivetermit nähdä niiden osittaisautokorrelaatiofunktioista [4]. Allaolevassa kuvassa (Kuva 32) on esimerkkinä pH:n osittaisautokorrelaatiofunktio, josta mukaan on otettu 5 viivetermiä.



Kuva 32: pH:n osittaisautokorrelaatiofunktio.

Otetaan mukaan seuraavat viipeet:

- DO_{t-1} , DO_{t-2} ja DO_{t-3}
- $Gluk_{t-1,2}$
- $Lakt_{t-1,\dots,7}$
- $Ig_{t-1,2}$
- $pH_{t-1,\dots,5}$

Neuroverkkoon pitää mennä jokaisella kierroksella jokainen sisäänmeno, eli toisin sanoen kunkin aikasarjan tulee olla yhtä pitkiä. Koska viiveaikasarjat olivat viipeen kertaluvun verran lyhyempiä kuin alkuperäiset aikasarjat, piti näitä lyhentää suurimman esiintyvän viipeen verran, eli nyt 7.

Lopuksi systeemin sisääntulot piti vielä normalisoida, sillä tämän on huomattu parantavan neuroverkkojen suorituskykyä [2]. Normalisointi tehtiin kaavan (3) mukaan, jossa x_i on mitattu arvo, x_{\min} ja x_{\max} kunkin ajon pienin sekä suurin mitattu arvo, ja y_i siis normalisoitu arvo s.e. $y_i \in [0,1]$.

$$y_i = (x_i - x_{\min}) / (x_{\max} - x_{\min}) \quad (3)$$

Ennen kuin verkkoa voidaan alkaa opettaa, tulisi vielä päättää sen muoto. Käytännössä tämä toteutetaan kuitenkin simuloimalla verkkoa erilaisilla konfiguraatioilla, kuten muuttamalla neuronien siirtofunktioita

sekä piilokerrosten että niissä esiintyvien neuronien lukumäärää.

Neuronien lukumäärän lähtökohtaa varten on olemassa Baum-Hausslerin sääntö (kaava 4) [4], mutta tämä on vain lähtökohta ja vasta opettamalla sekä testaamalla opetetun verkon toimintaa päästään mahdollisimman hyvään lopputulokseen.

$$N_n \leq \frac{N_{train} E_{tolerance}}{N_{pts} + N_{output}} \quad (4)$$

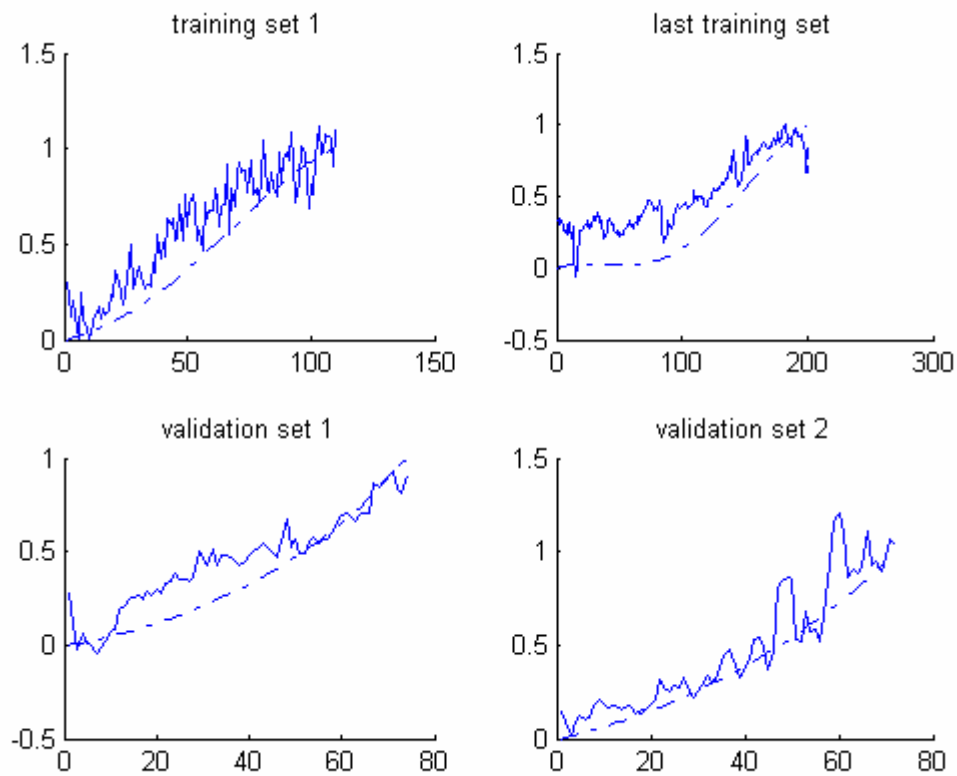
N_n on neuronien lukumäärä, N_{train} on opetusnäytteiden lukumäärä, $E_{tolerance}$ haluttu virheen toleranssi, N_{pts} kunkin opetusnäytteen sisältämä sisäänmenojen määrä, sekä N_{output} ulostulojen lukumäärä.

Nyt opetusdataa oli käytössä 674 kappaletta (kaksi ajoa on varattu mallin todentamiseen) ja mikäli virheen toleranssi haluttaisiin pitää varsin huonossa kokoluokassa 10^{-1} , saadaan neuronien maksimilukumääräksi 2,5. Opetusdataa ei siis ollut ainakaan liikaa, sillä näinkin monimutkainen malli vaatii hyvin toimiakseen huomattavasti enemmän (luokkaa 10 – 20 [5]) neuroneita. Koska opetusdataa ei ollut mahdollista saada lisää, ainoa vaihtoehto oli vähentää verkon sisäänmenojen määrää.

Baum-Hausslerin kaava on kuitenkin vain peukalosääntö, joten sivuutetaan se toistaiseksi ja palataan siihen tulosten analysoinnissa.

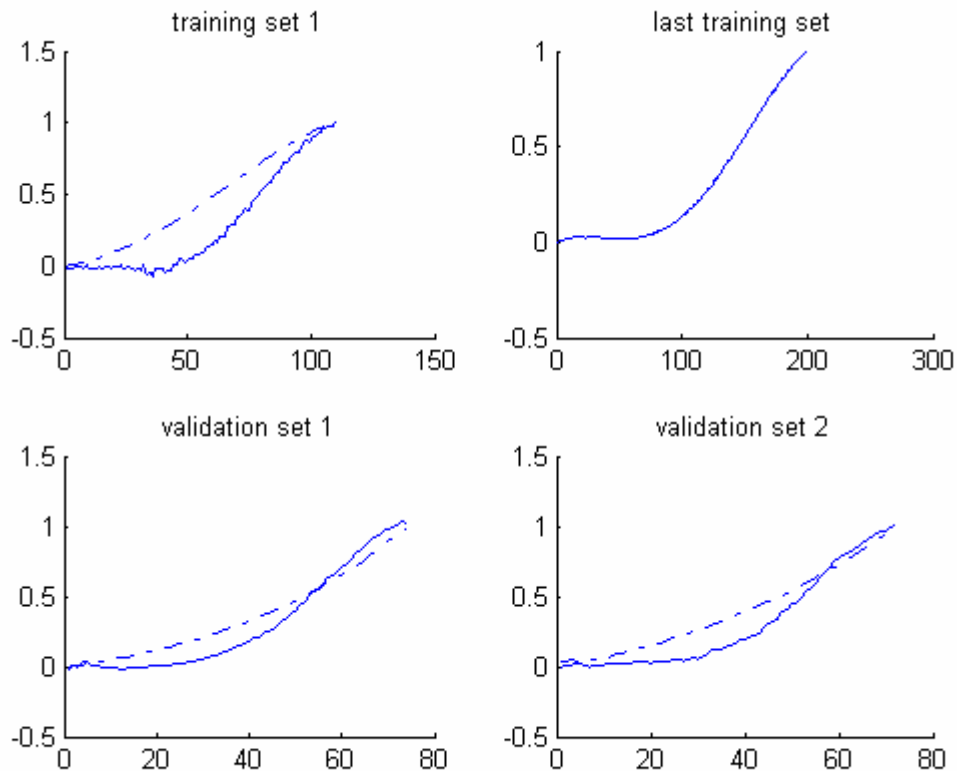
4.5.3. Opetus ja simulointi

Liikkeelle lähdettiin 10 hyperbolisella siirtofunktiolla varustetulla neuronilla yhdessä piilokerroksessa. Ulostulokerroksen ainoa neuron oli lineaarinen. Tällä konfiguraatiolla ei päästy hyviin tuloksiin, kuten oheisesta kuvasta (Kuva 33) nähdään. Kaikki kyseisen kuvan käyrät on piirretty opetuksen jälkeen, eli valmista verkkoa on simuloitu sekä kahdella opetuksessa käytetyllä ajolla että kahdella validointiin jätetyllä ajolla. Kuvissa katkoviiva kuvaa kohdefunktiota, yhtenäinen viiva verkkoa simuloimalla saatuja tuloksia.



Kuva 33: Opetetun verkon antamat simulaatiotulokset. Verkossa oli 10 neuronia yhdessä piilokerroksessa, virheen haluttu toleranssi 10^{-1} .

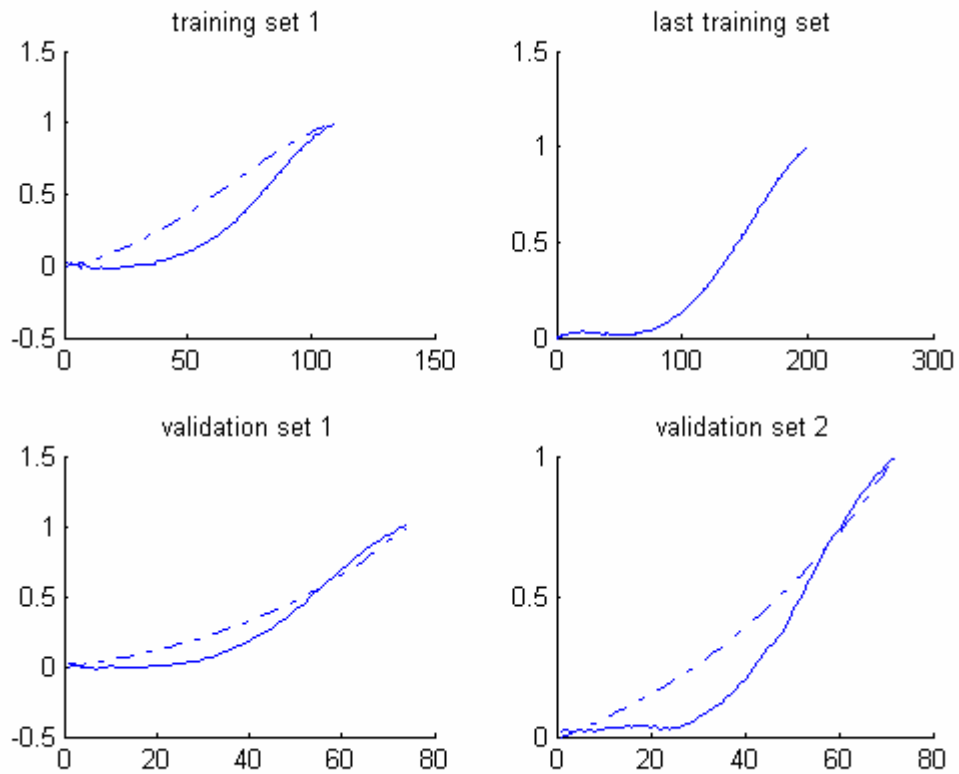
Simuloimalla verkon toimintaa myös opetusajoilla nähdään selvästi, että verkon opetus on jäänyt kesken. Tämä johtuu liian suuresta virheen toleranssista, johon päästiin jo vain muutaman opetusepookin jälkeen. Toleranssia pienennettiin neljällä dekadilla, eli arvoon 10^{-5} , vaikka Baum-Hausslerin säännön mukaan tämä edelleen alentaa mahdollisten neuronien määrää.



Kuva 34: Pienemmällä virheen toleranssilla opetetun verkon simuloinnit kahdella opetusajolla sekä kahdella validointiajolla.

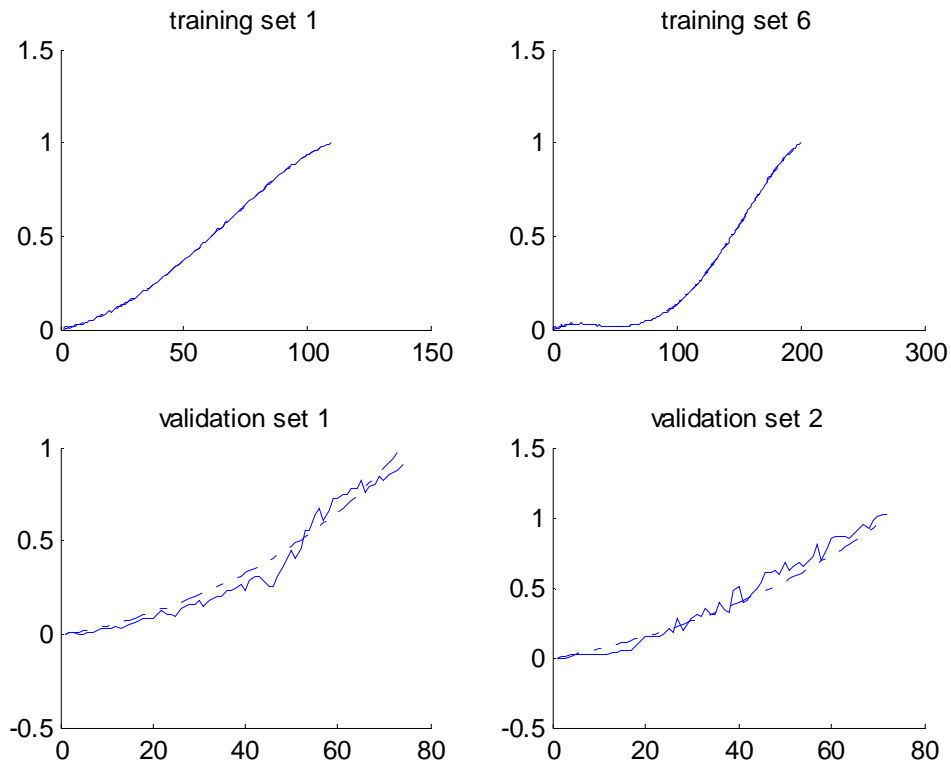
Virheen toleranssin pienentäminen paransi mallin toimintaa huomattavasti. Ongelmaksi muodostui kuitenkin se, että verkko oppi liian hyvin eri ajokerrat, jolloin mallin kyky yleistää huonontuu. Tämä näkyy eritoten opetusdatan simuloinneista (Kuva 34); verkko osaa mallintaa viimeisen opetukseen käytetyn ajon lähes täydellisesti, kun taas ensimmäisen opetusajon simulointi ei onnistu enää niin hyvin. Lisäksi tämän voi huomata viimeisen opetusajon kohdefunktion muodon toistumisessa kaikissa simuloinneissa.

Verkon kykyä yleistää voitaisiin parantaa kasvattamalla virheen toleranssia, mutta nyt se ei kuitenkaan ole suotavaa, sillä se saa mallin ulostulon todella epätasaiseksi (kuten Kuva 35 osoittaa). Yleistäminen oli siis tehtävä vähentämällä neuronien lukumäärää. Liian montaa neuroniam ei verkosta voi kuitenkaan poistaa, sillä tällöin sen kyky mallintaa näinkin monimutkaista systeemiä huonontuu liiaksi. Käytännössä tämä näkyy siten, että verkon opettamisvaiheessa ei ikinä saavuteta haluttua virheen toleranssia. Nyt tämä raja tuli vastaan 4 neuronilla, eli mallissa tulee olla vähintään 5 neuroniam. Neuronien jakaminen kahteen tai useampaan piilokerrokseen ei myöskään vaikuttanut sen toimintaan merkittävästi. Kuten kuvasta (Kuva 35) nähdään, verkon yleistyskyky ei juurikaan parantunut neuronien määrän vähentämisellä.



Kuva 35: Opetetun verkon simuloidut tulokset, kun sen piilokerroksen neuronien määrä on pudotettu viiteen.

Vielä eräs mahdollisuus parantaa mallin yleistämiskykyä oli yhdistää kunkin ajon tiedot ja tehdä niistä yksi iso opetussarja järjestämällä ne uudelleen ajan suhteen. Tämä mahdollisesti eliminoi verkon mukautumisen vain yhteen, viimeisenä käytettyyn opetussarjaan.



Kuva 36: Yhdistetyillä ajoilla opetettu verkko sekä sillä simuloitut ajot. Training set 1 sekä 6 ovat osa opetusdataa.

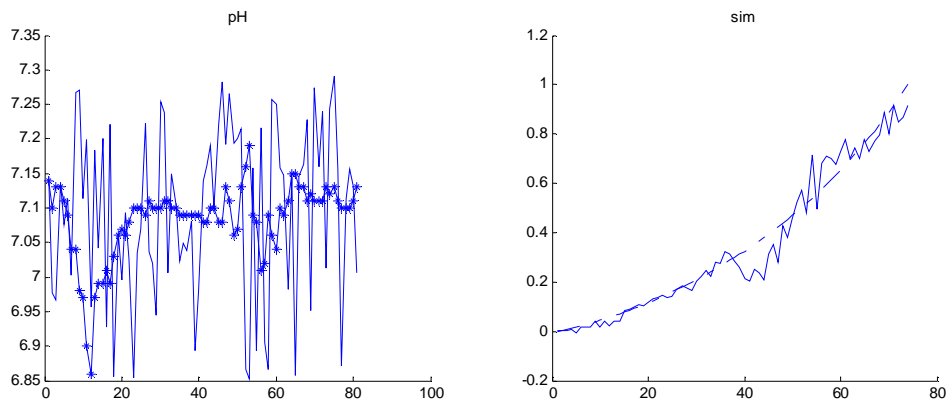
Kuten oheisesta kuvasta (Kuva 36) nähdään, opetukseen käytettyjen ajojen yhdistäminen paransi verkon suorituskykyä. Tähän pääsemiseksi verkkoon piti kuitenkin lisätä toinen piilokerros sekä neuroneita. Ensimmäinen piilokerros koostui 13 neuronista, toinen kahdeksasta. Kummankin kerroksen neuronien siirtofunktiona toimii hyperbolinen tangentti.

Vaikka verkon suorituskyky parani, todentamiseen jätettyjen ajojen simuloinnissa on vielä toivomisen varaa.

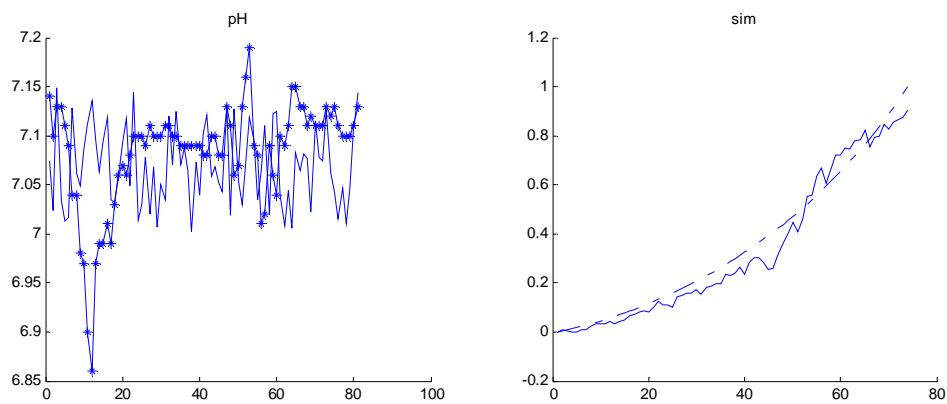
Mallia voidaan kuitenkin vielä yksinkertaistaa. Sisäänmenojen lukumäärä on varsin suuri, joten kokeiltiin, miten viivetermien vähentäminen vaikuttaa mallin suorituskykyyn.

Monen iteraation jälkeen huomattiin, että mallin suorituskyky ei laskenut merkittävästi (ks. liite 3, kuva 1), vaikka siitä jättää pois viipeet DO_{t-3} , $Gluk_{t-2}$, $Lakt_{t-5,6,7}$ sekä $pH_{t-4,5}$.

Kokeiltiin vielä, miten edellä opetetun mallin antama ennuste muuttuu, kun mitatuissa suureissa tapahtuu muutoksia. Tämän avulla toivoimme löytävämme suureille raja-arvoja sekä niiden perusteella annettavia suosituksia.



Kuva 37: Oikealla mallin ennuste (yhtenäinen viiva, katkoviiva on mitattu tuotto), kun validointiin käytetyn ajon mitattu pH-arvo on muutettu satunnaisesti välillä [6.85, 7.3] (yhtenäinen viiva). Ajon oikeat pH-arvot on merkitty tähdillä.



Kuva 38: Mallin ennuste, kun pH-arvon on annettu vaihdella satunnaisesti välillä [7.0, 7.15].

Oheisesta kuvasta (Kuva 37) nähdään, että pH:n suuri vaihtelu saa mallin ennusteen käyttäytymään rauhattomasti. Sama ilmiö on havaittavissa oheisesta kuvasta (Kuva 38), jossa pH on satunnaista kohinaa suppeammalla välillä. Suurempaa vaikutusta simuloituun tuottoon sillä ei kuitenkaan ole, joten mallin hyvyys voidaan kyseenalaistaa.

Näiden tulosten valossa voidaan sanoa, että huolimatta käytettävissä olevan datan samankaltaisuudesta (sama reaktori, sama solulinja), jokainen ajokerta on kuitenkin niin yksilöllinen, että yleisen mallin luominen on hankalaa. Tähän vaikuttaa varmasti ajoissa tapahtuneet ulkopuoliset tekijät, kuten inhimilliset virheet tai esimerkiksi kuolleiden solujen poistaminen, joita ei saada malliin mukaan. Myös mittaustiedon reikäisyys, sen täyttäminen sekä sen vähyys Baum-Hausslerin säännön mukaan vaikuttavat varmasti tuloksiin.

Jatkotoimenpiteinä voisi neuroverkkojen osalta olla kokeilla myös muunlaisia verkkoja. Myös käytettävissä olevaa dataa voitaisiin käyttää erilailla: yhdistää kaikki ajot, järjestää ne ajan suhteen ja poimia testaamiseen käytettävät mittaukset tästä joukosta. Myös kohdefunktiota voisi vaihtaa, vaikkapa hetkelliseksi tuotoksi.

5. Johtopäätökset

Bioreaktorin dynamiikkaa ei saatu kuvattua sillä tarkkuudella, että jonkinlaisen optimin hakeminen ohjaukselle olisi ollut järkevää tai mahdollista. Viitteellisiä hälytysrajoja eri aineiden pitoisuuksille tai pH-arvolle ei myöskään kyetty projektin puitteissa haarukoimaan. Tehdyt havainnot ovat silti mielenkiintoisia, ja tarjoavat lisätietoja tuotantoprosessista ja sen onnistumiseen vaikuttavista tekijöistä. Edelleen nyt muodostetut aikasarjamallit sekä muut tehdyt havainnot antavat viitteitä mahdollisen jatkotutkimuksen varalle, ja ennustemallilla voidaan tietyissä rajoissa tutkia simuloimalla, miten eri muuttujien muutokset kokonaisuuteen vaikuttaisivat.

Käytetyn aineiston laadusta (reikien määrä, ajojen pituus, ajojen vertailukelpoisuus) johtuen tehtävät johtopäätökset on kuitenkin varustettava melko suurella varauksella, ja käyttökelpoisimmat havainnot ovatkin varmasti eniten viitteellisiä. Ajojen vertailukelpoisuudesta ei aineiston vähäisyyden vuoksi lopulta juuri välitetty, mikä hankaloittaa tulosten tulkitsemista. Myöskään reikien täyttämisen vaikutusta tutkimustuloksiin ei ehditty tutkia. On mahdollista, että käyttämällä lineaarisella interpolaatiolla täytettyä tuotantoaineistoa estimoitujen mallien kertoimet voisivat poiketa paljonkin nyt saaduista.

Käytetyistä menetelmistä lineaarinen regressio onnistui tyydyttävästi, tuottaen tulokseksi reaktorin toimintaa selittävän stokastisen differenssiyhtälön. Myös kuvaajia ja samankaltaisten ajojen eroja tutkimalla saadut havainnot tarjoavat selityksen tiettyjen ajojen epäonnistumiselle. Näitä havaintoja ei kuitenkaan voida ilman jatkotutkimusta yleistää koskemaan bioreaktorin toimintaa yleisesti. Neuroverkot eivät tällä kertaa tuottaneet käyttökelpoista mallia, mikä johtui ilmeisesti suurimmalta osin menetelmän vaatimasta datamäärästä.

5.1. Havainnot järjestelmän dynamiikasta

Lineaarinen regressio ja ajojen silmämääräinen tarkastelu tuottivat joitakin samansuuntaisia tuloksia, ja niille onkin syytä antaa suurin painoarvo reaktorin käyttäytymistä selittävinä tekijöinä. pH-arvon poikkeama tasapainoarvosta 7,0 näyttäisi korreloivan vasta-aineen vuorokausituoton kanssa, ja myös regressiomallissa ajojen erotuksille pH nousee muita selvästi merkitsevämmäksi selittäjäksi. pH-arvon merkitykseen vaikuttaa varmasti myös se, että se reagoi joidenkin muiden aineiden pitoisuuksien muutoksiin. Näin itse taustaprosessia pitäisi tuntea paremmin, jotta pH:n ja tiettyjen aineiden pitoisuuksien yhteisvaikutuksesta voitaisiin tehdä tarkempia johtopäätöksiä.

pH-arvon lisäksi hapen osapaine oli tavallisen lineaarisen regression perusteella merkittävin vasta-ainekonsentraation selittäjä, ja sillä näyttäisi olevan merkitystä myös samankaltaisten ajojen tutkimuksen perusteella. Stokastisen differenssiyhtälön perusteella hapen osapaineen kasvu tosin vaikuttaisi positiivisesti vasta-ainepitoisuuteen. Kuitenkin johtuen siitä, että tässä mallissa hyvin suuri osa tiedosta on koodattu ensimmäiseen muuttujaan, eli vasta-ainepitoisuuteen viiveellä 1, ei mallin kertoimista voida tehdä suuria johtopäätöksiä eri muuttujien merkityksestä.

è pH-arvon poikkeamat kauas tasapainoarvosta 7,0 näyttäisivät vaikuttavan vasta-ainetuotannon

heikentymiseen.

è Myös hapen osapaineen suurella arvolla näyttäisi olevan vasta-aineen tuottoa heikentävä vaikutus.

Pelkästään silmämääräisesti tehtiin lisäksi muutamia perusteltuja havaintoja siitä, mitkä tekijät vaikuttivat tarkasteltujen ajojen vertailukohtaa huonompaan onnistumiseen. Kuvaajien perusteella ammoniumioni- ja glutamiinipitoisuuksilla voisi olla merkitystä ajojen onnistumiselle; molempien osalta liian suureksi kasvavat pitoisuudet näyttäisivät haittaavan vasta-aineen tuotantoa. Näistä ammoniumionin merkitystä myös tavallinen lineaarinen regressio tuki heikosti. Havaintoja ei kuitenkaan voida yleistää koskemaan kaikkia ajoja vastaavalla tavalla, kuin pH:n ja hapen osapaineen osalta.

è Edellä mainittuja muuttujia hieman heikompi vaikutus tuntuisi olevan ammoniumioni- ja glutamiinipitoisuuksilla. Molemmissa liian korkea pitoisuus vaikuttaa vasta-ainetuotantoa heikentävällä tavalla.

Myös voimakkaat muutokset pumppujen kiertonopeuksissa näyttäisivät johtavan helpommin tuotantoajon epäonnistumiseen verrattuna ajoihin, joissa pumppuja oli säädetty rauhallisemmin. Tämä voi johtua esimerkiksi edellä mainittujen aineiden pitoisuuksien merkityksestä vasta-aineen tuotannolle.

Erityisesti lineaarisen regression avulla saadut tulokset antavat suuntaa siihen, mihin mahdollisia myöhempiä tutkimuksia kannattaisi kohdistaa. Ristitermit *DOxLakt* ja *GlukxLakt* näyttäisivät vaikuttavan vasta-ainepitoisuuteen, joten niiden suuruuksia kannattaisi tutkia edelleen ajojen hyvyyden suhteen. Menetelmäksi sopii esimerkiksi näiden arvojen tulostaminen *xy*-kuvaajaan, josta voi edelleen hakea yhtenäisiä alueita ajojen hyvyyden mukaan.

Lisäksi muodostetun stokastisen differenssimallin avulla voidaan simuloida jo suoritettujen ajojen käyttäytymistä, kun jokin muuttuja saa alkuperäisestä poikkeavia arvoja. Esimerkiksi jonkin aineen pitoisuuteen voisi lisätä nousevan tai laskevan trendin, tai voimakasta kohinaa, ja tutkia, mihin suuntaan vasta-aineen pitoisuudet kehittyvät. Koska muodostettu malli selittää tutkittuja ajoja melko hyvin, voisi tällä tavalla saavuttaa ainakin suuntaa antavia tuloksia esimerkiksi muuttujien kriittisistä hälytysrajoista.

Simuloinnin lisäksi differenssimalli sopii hyvin varovaiseen ennustamiseen jo aloitettujen ajojen osalta, jos mallissa esiintyvien muuttujien kehittyminen osataan ennakoida. Ennustamisen aikahorisontti on korkeintaan 10 päivää, jonka jälkeen ennusteista lähes varmuudella tulee hyvin virheellisiä.

è Jatkossa tuotantoprosessin tutkimuksessa kannattaisi keskittyä stokastisessa differenssiyhtälössä esiintyvien ristitermien hyvien alueiden löytämiseen. Näiden muuttujien vaikutusta vasta-ainetuotantoon voi tutkia myös simuloimalla muodostetun mallin avulla.

è Jo aloitettujen ajojen vasta-ainetuotannon kehittymistä seuraavien päivien aikana voi varovaisesti ennustaa stokastisen differenssimallin avulla.

5.2. Suositukset tuotantoprosessin hoitamiseksi

Silmämääräisesti tehtyjen havaintojen perusteella onnistuneiden ja epäonnistuneiden ajojen välillä oli

usein se ero, että onnistuneessa ajossa pumppujen nopeuksia oli säädetty tasaisemmin. Myös pH:n suuren varianssin todettiin vaikuttavan todennäköisesti vasta-ainetuottoa heikentävä tavalla. Koska pH:ta säädetään täysin automaattisesti, on mahdollista, että aineita solutilaan tuovien ja sieltä poistavien pumppujen nopeuden nopea muutos aiheuttaa pH:ssa lyhytaikaisen häiriön, johon säädin ei kykene reagoimaan.

- è Pumppuja tulisi säätää mahdollisimman tasaisesti, välttämällä suuria äkillisiä muutoksia. Esimerkiksi kaksi kertaa viisi yksikköä lisää nopeutta on parempi muutos kuin kerran kymmenen yksikköä.

Tämän lisäksi pH-arvoa kannattaisi seurata tarkasti, jotta poikkeaman arvosta 7,0 ollessa epätavallisen suuri, voidaan tehdä korjaavia toimenpiteitä. Ajojen ennenaikaisen lopettamisen suosittamiseksi eri tilanteissa tarvittaisiin kuitenkin tarkempaa tutkimusta ajojen keskimääräisestä sujumisesta ja tuotannon kustannuksista.

5.3. Suositukset operatiivisen toiminnan kehittämiseksi

Käytettävissä olleen datan laatu oli melko heikkoa tutkimuksen kannalta. Havaintopisteiden määrä suhteessa aikasarjan pituuteen sekä se, että eri ajoista oli kirjattu ylös eri muuttujien arvoja, hankaloittivat tutkimustyötä vieden aikaa ja huomiota varsinaiselta analyysiltä. Lisäksi tulosten tulkittavuus on huomattavan heikko verrattuna siihen, että käytettävissä oleva aineisto olisi ollut likimain ehjää. Myös osan mittaustiedosta kerääminen paperilta vei hieman turhan paljon aikaa.

Edellytyksiä tuotantoprosessin tutkimiselle ja optimoinnille voisi parantaa helposti ottamalla käyttöön mittaustiedon ensisijaisesti sähköisen dokumentoinnin. Tämän lisäksi aineistoa voi säilyttää myös paperilla, jos sähköisen arkiston pysyvyys ei ole riittävä. Sähköisen dokumentoinnin käyttämisestä ei kuitenkaan tule kuin mahdollisesti pieni aloituskustannus (tietokoneen hankinta), mutta sen suomat hyödyt olisivat tutkimustyön kannalta erittäin arvokkaita.

Tämän lisäksi tutkimusta helpottaisi datan keräämisen syklin tihentäminen. Mittaustietoa tulisi kerätä vähintään joka arkipäivä, ja useamminkin jos mahdollista. Myös viikonloput ovat tärkeitä tutkimuksen kannalta. Lisäksi samalla voisi miettiä, miten epänormaalit tapahtumat ja niiden vaikutus prosessille dokumentoidaan, jotta tietoa voidaan käyttää hyväksi mallintamisessa. Tärkeintä on kuitenkin datamäärä yleisesti; ajojen pituuden kannattaisi olla noin 200-300 havaintoa, ja reikien määrä siitä korkeintaan 5 %.

Datan keräämiseen vaikuttavat tietysti mittausten työläys ja sitä myötä tulevat kustannukset, sekä toisella puolella tuotantoprosessin tarkemmasta tutkimisesta mahdollisesti saatavat hyödyt.

- è Mittausdata kannattaisi kerätä ensisijaisesti sähköiseen muotoon, jotta tuotantoprosessin tutkimiselle jatkossa on hyvät edellytykset.
- è Jotta käytetyillä menetelmillä saataisiin varmasti luotettavia tuloksia, tulisi mittausten sykliä tihentää vähintään yhteen mittaukseen per päivä.

6. Viitteet

- [1] Banga, J., Balsa-Canto, E., Moles, G., Alonso, A. (2002) *Dynamic Optimization of Bioreactors: A Review*, Proc. Ind. Natn. Science Academy, 69A(3-4):257-265
- [2] Tholudur, A. & Ramirez, F. (1996) *Optimization of Fed-Batch Bioreactors Using Neural Network Parameter Function Models*, Biotechnol. Prog. 12 (3), s. 302-309, 1996
- [3] Shumway, R. & Stoffer, D. (2000) *Time Series Analysis and Its Applications*, s. 15. Springer
- [4] *NCSS help file*, NCSS 2000 -ohjelmisto, julkaisija NCSS, USA
- [5] Keinänen, A. (2003) *Tilastolliset menetelmät rikostaloustieteessä*, Joensuun yliopisto, Taloustieteiden laitos (www.joensuu.fi/taloustieteet/ott/opinnaytteet/lisurit/keinanen.pdf)
- [6] Greene (1997). Greene, W. H (1997) *Econometric analysis*, 3rd edition, s. 420, Prentice-Hall, inc.
- [7] Ilkka Mellin: Aikasarja-analyysi: Dynaamiset regressiomallit (<http://www.sal.tkk.fi/Opinnot/Mat-2.128/IMLUennot06/EADY100.pdf>.)
- [8] Feng Ling, Zing Huo Yu, Shirley Gregor and Richard Irons. (1995) *Time Series Forecasting with Neural Networks*. Complexity International. Vol. 2.
- [9] Pigram, G.M., MacDonald, T.R. (2001) *Use of Neural Network Models to Predict Industrial Bioreactor Effluent Quality*. Environmental Science & Technology. Vol. 35. pp. 157 – 162.

7. Liitteet

- [1] Laktaattipitoisuuden muutoksen ja vuorokausituoton ristikorrelaatiot
- [2] pH-arvon poikkeaman tasapainosta ja vuorokausituoton ristikorrelaatiot
- [3] Neuroverkon toiminta karsitulla sisäänmenojoukolla
- [4] Yhteenveto projektin sujumisesta