



Osaamista  
ja oivallusta  
tulevaisuuden  
tekemiseen

Simo Mäkelä

# Oksimetrian soveltuvuus aivoverenkierron tutkimiseen ensihoidossa

Metropolia Ammattikorkeakoulu

Insinööri (AMK)

Tieto- ja viestintätekniikka

Insinööryö

5.2.2021

Tekijä Otsikko Sivumäärä Aika	Simo Mäkelä Oksimetrian soveltuvuus aivoverenkierron tutkimiseen ensihoidossa 41 sivua + 2 liitettä 5.2.2021
Tutkinto	Insinööri (AMK)
Tutkinto-ohjelma	Tieto- ja viestintäteknikka
Ammatillinen pääaine	Hyvinvointi- ja terveysteknologia
Ohjaajat	dosentti Jouni Nurmi, FinnHEMS Oy lehtori Sakari Lukkarinen
<p>Opinnäytetyössä suunniteltiin ja toteutettiin data-analyysiprosessi ja siihen liittyvä ohjelmakoodi kliinisen tutkimuksen (BOPRA) tarpeisiin. Työssä käytettiin aikasarja-analyysiä fysiologisten signaalien välisen yhteyden tutkimisessa. Tutkittavat signaalit olivat keskivaltimopaine ja aivokudoksen happikyllästeisyys. Työn tavoitteena oli tuottaa tietoa anestesiassa olevien potilaiden aivoverenkierrosta ja mittausmenetelmän soveltuvuudesta ensihoidon toimintaympäristöön.</p> <p>Datan luonteen hahmottamisessa noudatettiin eksploraatiivisen data-analyysin ajatusmallia, jossa graafisten ja tilastollisten yhteenvetojen avulla muodostetaan johtopäätöksiä aineistosta. Laskennassa hyödynnettiin aiemmissa tutkimuksissa kuvailtuja menetelmiä, joiden soveltuvuutta nyt päästään arvioimaan BOPRA-tutkimuksessa ensihoidon viitekehyksessä. Aiemmat tutkimukset keskittyivät pääasiassa tehohoitoon, josta ensihoidon toimintaympäristö poikkeaa pääasiassa liikehäiriöiden ja epästabiilin alkutilanteen tuomien vaihteluiden määrässä sekä potilaskontaktin lyhydessä.</p> <p>Ohjelmallisessa toteutuksessa käytettiin työkaluina Python-ohjelmointikielelle kehitetyjä data-analyysin työkaluja, kuten NumPyä, pandasia ja matplotlibiä. Ohjelmakoodin kirjoittamisessa käytettiin Jupyter notebook -ympäristöä sekä tavallisia Python-skriptejä ja -moduuleja. Suoritusympäristön hallintaan käytettiin conda-ohjelmistoa. Versionhallintajärjestelmänä toimi git. Ohjelmakoodin ja teknisen dokumentaation julkaisualustana hyödynnettiin GitHub-palvelua. Väliraportointiin kehitettiin oma järjestelmä hyödyntäen Markdown-merkintäkieltä ja PHP-verkkosivuja.</p> <p>Tuloksena syntyi tutkimuskäyttöön soveltuva työkalu BOPRA-datan käsittelyyn ja raportointiin. Kehitysprosessin kuluessa tutkijat saivat uudenlaisia välineitä tutkimusdatan tarkasteluun, ja tekijästä tuli osa tutkimustiimiä.</p>	
Avainsanat	data-analyysi, aivoverenkierto, rSO <sub>2</sub> , COx, NIRS

Author Title Number of Pages Date	Simo Mäkelä Applicability of Tissue Oximetry for Assessment of Cerebral Circulation in Prehospital Care 41 pages + 2 appendices 5 February 2021
Degree	Bachelor of Engineering
Degree Programme	Information Technology and Communications
Professional Major	Health Technology
Instructors	Jouni Nurmi, Adjunct Professor, FinnHEMS Ltd. Sakari Lukkarinen, Senior Lecturer
<p>This paper describes the design and implementation of a data analysis process for a clinical study (BOPRA), with its associated program code. Time series analysis was used in the examination of association between physiological signals. The signals under inspection were the mean arterial pressure and cerebral tissue oxygen saturation. The objective was to generate knowledge on cerebral circulation of patients under anesthesia and the applicability of the measurement method in a prehospital environment.</p> <p>The concept of explorative data analysis (EDA) was employed to form conclusions of data characteristics by using graphical representations and summary statistics. The computations rely on methods already described in research, the applicability of which can now be scrutinized in a prehospital context by the BOPRA study. Past research has mostly been conducted in an intensive care setting, wherefrom the prehospital environment differs notably by the number of fluctuations caused by movement and initial patient instability, as well as the shorter duration of patient contact.</p> <p>The program code was implemented using the Python data analysis ecosystem, comprising tools such as NumPy, pandas and matplotlib. The code was written into Jupyter notebooks and plain Python modules and scripts. Conda was used for environment management and git was used for version control. The publishing platform for source code and technical documentation was GitHub. For intermediate reporting, a custom implementation was created using Markdown and PHP web pages.</p> <p>In this work, a software tool adequate for scientific use was created to process and report BOPRA data. During the development process, the researchers were presented new means to examine the data, and the author became part of the research team.</p>	
Keywords	Data analysis, Cerebral circulation, rSO <sub>2</sub> , COx, NIRS

# Sisällys

Lyhenteet ja käsitteet

1	Johdanto	1
2	Aivoverenkierto ja sen mittaaminen	2
2.1	Autoregulaatio	2
2.2	Mittausmenetelmät	5
2.2.1	Valtimopaine	5
2.2.2	Aivojen oksimetria	6
2.3	Autoregulaatiota kuvaavat indeksit	8
3	Vaatimusmäärittely ja kehitysprosessi	11
4	Laadulliset tavoitteet	13
4.1	Validoitavuus ja läpinäkyvyys	13
4.2	Toistettavuus ja uusittavuus	14
4.3	Tietosuoja ja eettisyys	15
5	Toteutus	16
5.1	Työkalut	16
5.2	Suoritusympäristö	17
5.3	Visualisointi	20
5.4	Datan esikäsittely	22
5.4.1	Raakadatan jäsentäminen ja lataaminen	23
5.4.2	Aaltomuotojen tunnistaminen	23
5.4.3	Aikasarjojen tasoittaminen	24
5.4.4	Aikasarjojen kohdistaminen	27
5.4.5	Virheellisten arvojen poisto	29
5.5	COx-indeksin laskenta	30
5.6	Valtimopaineen optimointi	32
5.7	Viimeistely visualisointi	34
6	Pohdinta	36
	Lähteet	38

Liitteet

Liite 1 Tekninen dokumentaatio

Liite 2 Data-analyysin dokumentaatio

## Lyhenteet ja käsitteet

<b>autoregulaatio</b>	<i>Itsesääätely.</i> Tässä työssä tarkoitetaan aivojen kykyä säädellä omaa verivirtaustaan: engl. <i>cerebral autoregulation</i> .
<b>BOPRA</b>	<i>Brain Oxygenation during Prehospital Anesthesia.</i> Aivojen happautumista käsittelevä tutkimus [1].
<b>CBF</b>	<i>Cerebral Blood Flow.</i> Aivoihin virtaava verimäärä aikayksikköä kohden. Aikuisella tyypillisesti noin 750 ml/min.
<b>COx</b>	<i>Cerebral Oximetry Index.</i> Autoregulaatiota kuvaava arvo, joka on valtimopaineen (MAP) ja aivokudoksen happisaturaation (rSO <sub>2</sub> ) välinen liukuva Pearsonin korrelaatiokerroin [2].
<b>CPP</b>	<i>Cerebral Perfusion Pressure.</i> Aivojen perfuusiopaine, joka on valtimopaineen (MAP) ja kallonsisäisen paineen (ICP) erotus.
<b>CSV</b>	<i>Comma-separated Values.</i> Taulukkomainen tiedostomuoto, jossa arvot ovat välimerkein toisistaan eroteltuina.
<b>CVR</b>	<i>Cerebral Vascular Resistance.</i> Aivoverisuonten virtausvastus.
<b>ICP</b>	<i>Intracranial Pressure.</i> Kallonsisäinen paine.
<b>iskemia</b>	Kudoksen paikallinen <i>verettömyys</i> .
<b>JSON</b>	<i>JavaScript Object Notation.</i> Avain-arvo -pareihin ja listoihin perustuva tiedostomuoto.
<b>MAP</b>	<i>Mean Arterial Pressure.</i> Verenpaineen keskiarvo valtimoissa.
<b>NIRS</b>	<i>Near-Infrared Spectroscopy.</i> Lähi-infrapunaspektroskopia.
<b>oksimetria</b>	<i>Happikyllästeisyyden</i> mittaust.
<b>pandas</b>	Python-ohjelmointikielellä käytettävä data-analyysikirjasto.
<b>perfuusio</b>	Veren <i>läpivirtaus</i> kudossmassaa ja aikayksikköä kohden. Aivoissa tyypillisesti noin 50 ml/100 g/min.
<b>rSO<sub>2</sub></b>	<i>Regional Oxygen Saturation.</i> Kudoksesta mitattu paikallinen happisaturaatioarvo.

## 1 Johdanto

Tässä opinnäytetyössä käsitellään BOPRA-nimisessä kliinisessä tutkimuksessa kerättyä dataa. BOPRA-tutkimuksessa (Brain oxygenation during prehospital anesthesia) tarkastellaan aivojen happeutumista ensihoidollisen anestesian aikana. Data koostuu ensihoitotehtävien aikana kerätyistä fysiologisista mitta-arvoista, joista oleellisimpana lähifrapunaspektroskopiaan (NIRS) perustuva aivokudoksen paikallinen happisaturaatio ( $rSO_2$ ). BOPRA-tutkimuksen tavoitteena on selvittää näistä mitta-arvoista muodostettujen parametrien yhteyttä kuolleisuuteen, sairastavuuteen ja elämänlaatuun [3]. Työn tilaaja on BOPRA-tutkimuksen järjestämisestä vastaavan lääkärihelikopteri-yhtiö FinnHEMS Oy:n tutkimus- ja kehitysyksikkö.

Eräs tutkittava parametri on ns. COx-indeksi, joka kuvaa aivoverenkierron tilaa, tarkemmin sen *autoregulaatiota*. Indeksien laskennassa käytetään syötearvoina keskivaltimopainetta (MAP) sekä edellä mainittua paikallista happisaturaatiota ( $rSO_2$ ).

Tämän työn ensisijainen tavoite on laatia ohjelmakoodi COx-indeksien määrittämiseksi annetusta datasta. Ohjelman tulee muodostaa tulosten validoimiseksi ja hyödyntämiseksi tarvittavat datatiedostot, raportit ja visualisoinnit. Toisena tavoitteena on tutkia, voidaanko BOPRA-datan mitta-arvoista tuottaa potentiaalisesti kliiniseen päätöksentekoon soveltuvaa tietoa.

Ongelmaa lähestytään eksploraatiivisen data-analyysin keinoin hyödyntäen signaalinkäsittelyn, aikasarja-analyysin ja fysiologian tuntemusta. Työssä tullaan kokeilemaan kuinka Aries ym. 2012 [4] kuvailema valtimopaineen tavoitetason estimointimenetelmä käyttäytyy ensihoidon toimintaympäristössä tuotetulla datalla.

## 2 Aivoverenkierto ja sen mittaaminen

Verenkierron tehtävä on tuoda kudoksille happea ja ravinteita sekä kuljettaa pois hiilidioksidia ja kuona-aineita. Toimiva verenkierto on solujen elämän edellytys. [5.] Sydämen pumppaustoiminnan tuottama paine-ero valtimoiden ja laskimoiden välillä aikaansaa veren virtauksen kudosten läpi. Elimistö kykenee vaikuttamaan tähän virtaukseen säätelemällä verisuoniston virtausvastusta suonten poikkipinta-alaa vaihtelemalla.

Aivojen sijainti kallon sisällä estää niiden vapaan laajentumisen. Kallontalon sisältämän materiaalin kokonaistilavuuden tulisi pysyä vakiona, jotta kallonsisäinen paine ei pääsisi nousemaan. Kallonsisäinen paine (ICP) rajoittaa aivoverenkiertoa ylläpitävää perfuusio-painetta (CPP):

$$CPP = MAP - ICP \quad (1)$$

jossa MAP on valtimopaineen keskiarvo, *mean arterial pressure*. Kallonsisäinen paine siis muodostaa valtimopaineelle vastapaineen, joten kallonsisäisen paineen kohotessa valtimopaineen tasolle, aivoverenkierto pysähtyy kokonaan.

Hapen tarve aivokudoksessa on sen aerobisen sokeriaineenvaihdunnan takia korkea. Ravinteiden saannin tulee olla keskeytyksetöntä, sillä aivoissa ei ole omia energiavarastoja. Tämä tekee aivoista hyvin alttiit *iskeemisille* vaurioille. Liiallinen verenvirtaus taas voi vaurioittaa aivokudosta, aiheuttaa turvotusta ja johtaa kallonsisäisen paineen nousuun. Aivoihin on näiden erityispiirteiden myötä kehittynyt verenvirtauksen tasaisena pitävä, *autoregulaatioksi* kutsuttu säätelymekanismi.

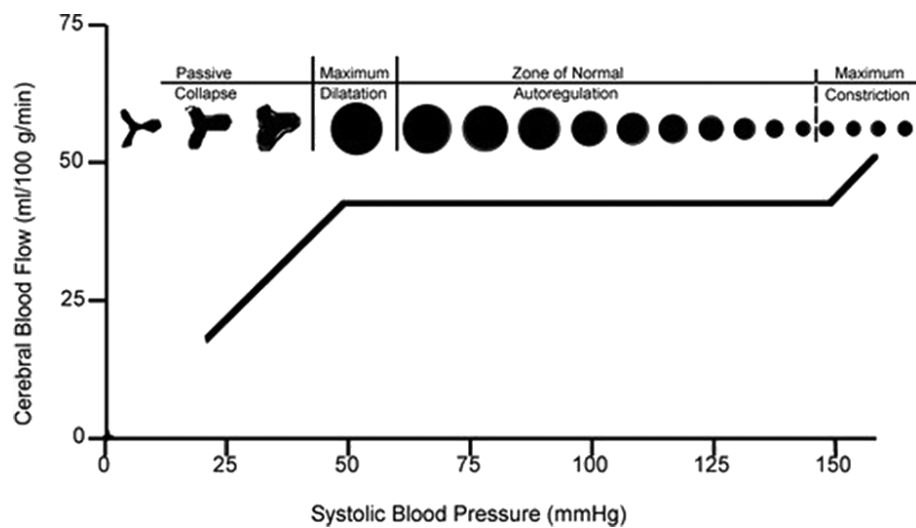
### 2.1 Autoregulaatio

Terveissä aivoissa verenvirtaus (CBF) pysyy autoregulaation ansiosta vakaana perfuusio-paineen vaihdellessa. Autoregulaatio kompensoi verenpaineen muutoksia säätelemällä aivoverisuoniston virtausvastusta (CVR). Aivojen verenvirtauksen, aivoverisuoniston vir-

tausvastuksen ja perfuusiopaineen välistä yhteyttä voidaan kuvata yhtälöllä 2:

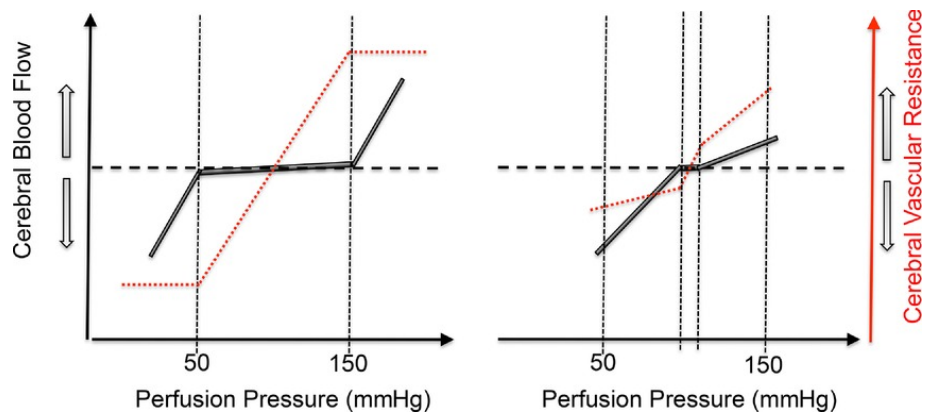
$$CBF = \frac{CPP}{CVR} \quad (2)$$

Autoregulaation toiminta-alueen alaraja on valtimopaine, jolla verisuonet ovat täysin laajentuneet. Ylärajalla taas suonet ovat supistuneet kapeimmilleen eivätkä enää paineen noustessa pysty rajoittamaan virtausta (kuva 1). Autoregulaatioalueen ulkopuolella aivo-verenvirtaus on passiivisesti riippuvainen systeemisestä valtimopaineesta.



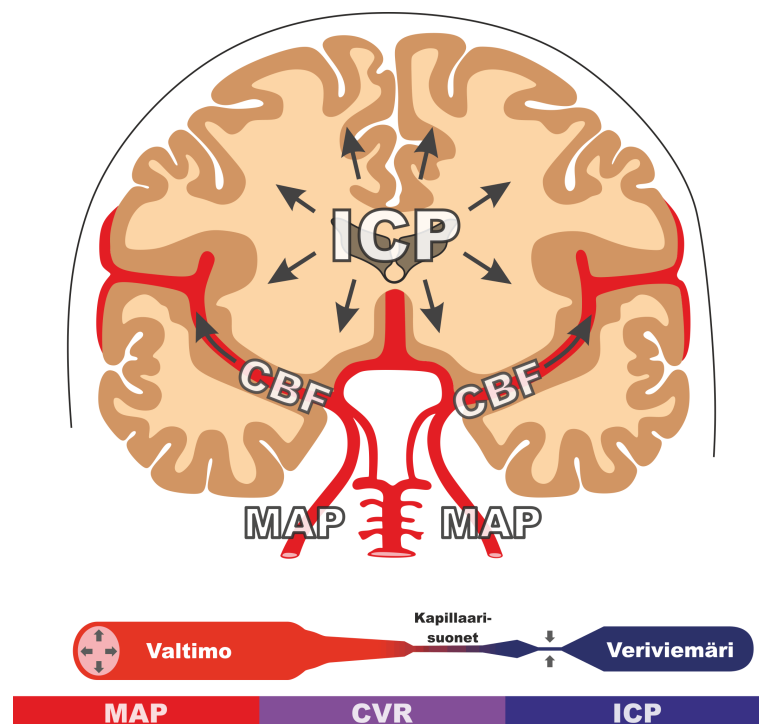
Kuva 1: Klassinen näkemys valtimopaineen ja aivoverenvirtauksen suhteesta (White, Venkatesh 2008 [6], Lassen 1959).

Autoregulaatioon vaikuttavat useat yksilölliset fysiologiset ja biokemialliset tekijät sekä jotkin sairaudet, kuten verenpainetauti, diabetes ja aivoverenkierron häiriöt [2]. Tuoreemman tutkimustiedon valossa autoregulaatioalue vaikuttaa tosiasiaassa olevan huomattavasti kuvan 1 esimerkkiä kapeampi. Ero on esitetty kuvassa 2.



Kuva 2: Oikealla tuoreempaan tutkimukseen perustuva luonnos autoregulaatiokäyrästä (Willie, Tzeng, Fisher, Ainslie 2014 [7]).

Autoregulaatio kykenee kompensoimaan kohonnutta perfuusiopainetta paremmin kuin alentunutta [7]. Kuvassa 2 on autoregulaatiokäyrän tasankovaiheen päihin merkitty pystysuuntaisilla katkoviivoilla autoregulaatioalueen ylä- ja alarajat. Näistä käytetään kirjallisuudessa lyhenteitä ULA ja LLA (engl. *Upper Limit of Autoregulation*, *Lower Limit of Autoregulation*). Kuvassa 3 on havainnollistettu aivojen verenvirtauksen määrittäviä tekijöitä ja aivoverenkierron anatomiaa.



Kuva 3: Aivojen verenvirtaukseen (CBF) vaikuttavia tekijöitä. Valtimopaine MAP on merkitty kuvaan *kaulavaltimoiden* päälle. Kallonsisäinen paine ICP muodostaa MAP:lle vastapaineen, joten aivoverenkiertoa ylläpitävä perfuusiopaine  $CPP = MAP - ICP$ . Kohonnut kallonsisäinen paine rajoittaa aivoverenkiertoa myös puristamalla suonia kasaan. Aivojen verenvirtaus CBF on merkitty kuvaan *keskimmäisen aivovaltimon* päälle. Autoregulaatio perustuu muuttuvaan aivoverisuoniston virtausvastukseen CVR. Kuva mukaeltu [5].

## 2.2 Mittausmenetelmät

Aivoverenvirtauksen (CBF) suoraan määrittämiseen ei ole olemassa jatkuvaa reaaliaikaista mittausmenetelmää, joka olisi käytännöllinen kliinisessä työssä [2]. Aivoverenkierroksen tilaa on tutkittu mm. Doppler-virtausmittauksin (laser, ultraääni), sekä välillisesti kallonsisäisen paineen vaihteluja mittaamalla [8].

Ensihoidon toimintaympäristössä ei kallonsisäistä painetta ole tällä hetkellä mahdollista mitata. Myöskään tarkan asettelun vaativa Doppler-mittaus ei ole käytännössä toteutettavissa. Ei-invasiivisen NIRS-oksimetrian tuottaman  $rSO_2$ -arvon on todettu olevan kelvollinen aivoverenvirtauksen (CBF) vaihteluja kuvaava korvikearvo eli surrogaatti [2].

### 2.2.1 Valtimopaine

Koska autoregulaatiota tutkittaessa mielenkiinnon kohteena on nimenomaan aivoverenvirtauksen painereaktiivisuus, tarvitaan jokin perfuusiopainetta (CPP) kuvaava mitta-arvo. Ilman invasiivista aivopainemittausta joudutaan CPP korvaamaan systeemisellä valtimopaineella (MAP). MAP:n katsotaan olevan kelvollinen korvike CPP:lle silloin, kun kallonsisäinen paine (ICP) ei ole kohonnut [2]. Kallonsisäisen paineen ollessa normaali ovat MAP ja CPP likimain samat. Kallonsisäisen paineen noustessa CPP laskee MAP:n alapuolelle yhtälön 1 mukaisesti. Tämä tulee huomioida mahdollisena virhelähteenä etenkin kallovammapotilaan mittaustuloksia tulkittaessa.

Tässä työssä MAP (*Mean Arterial Pressure*) eli valtimopaineen keskiarvo lasketaan valtimokanyylin kautta invasiivisesti mitatusta verenpainedatasta. Mittalaitteena tiedonkeruussa käytettiin ZOLL-monitori-defibrillaattoria (kuva 4).



Kuva 4: Ensihoidon käyttöön tarkoitettu ZOLL-monitori-defibrillaattori (zoll.com)

Ensihoidossa potilaan monitorointiin, 12-kytkentäisen EKG:n rekisteröintiin sekä sydämen defibrillointiin ja tahdistukseen käytetään kannettavaa, akkukäyttöistä monitori-defibrillaattoria. Kuvan 4 mukainen ZOLL-merkkinen laite on käytössä FinnHEMS-lääkärihelikopteryksiköissä, joissa BOPRA-tutkimuksen potilasdata on kerätty.

### 2.2.2 Aivojen oksimetria

Aivokudoksen paikallinen happikylläisyys ( $rSO_2$ ) voidaan määrittää lähi-infrapunaspekstroskopiaan (NIRS) perustuvalla menetelmällä, jossa mitataan infrapunavalon kudosläpäisevyyttä eri aallonpituuksilla [9]. Spektroskooppi määrittää oksija deoksihemoglobiinin suhteen kudoksessa *Beerin lakia* hyödyntäen [10]. Beerin laki kuvaa säteilyn absorboitumisen ja absorboivan aineen määrän välistä riippuvuutta [11].

NIRS-oksimetria poikkeaa yleisesti käytössä olevasta pulssioksimetriasta ( $SpO_2$ ) siten, että NIRS-oksimetri ei käytä *pulssiplotysmografiaa* valtimo- ja laskimopitoisuuksien eroteluun. Täten NIRS-oksimetrin lukema kuvaa kudoksen hapentarjonnan ja -kulutuksen välistä suhdetta, kun taas pulssioksimetrin lukema kertoo kudoksen hapensaannista. [9.]

NIRS-happikyllästeisyysmittaus kohdistuu valtimo- ja laskimovereen vastaavasti suhteessa 16 : 84. Tämä suhdeluku on kokeellisesti määritetty. [10.] NIRS-oksimetrian tuottaman saturaatiolukeman ( $rSO_2$ ) käyttö aivoverenvirtausta (CBF) kuvaavana mitta-arvona perustuu siihen, että virtauksen kiihtyessä laskimoihin päätyy enemmän hapekasta verta, mikä nostaa  $rSO_2$ -mittaustulosta.

Mittaustuloksen absoluuttiarvo vaihtelee potilaskohtaisesti yksilöllisten sirontaominaisuuksien ja mittauskentässä olevien kallon ulkopuolisten kudosten vaikutuksesta. Tämän epävarmuuden vuoksi absoluuttiarvosta ei voida tehdä johtopäätöksiä, joten NIRS-oksimetria soveltuu ainoastaan ajan myötä tapahtuvien muutosten havainnointiin. [10.] Otsalle liimattava NIRS-anturi on nopea ja helppo asettaa (kuva 5), joten se soveltuu käyttööminäisyyksiensä puolesta hyvin ensihoitoon.



Kuva 5: Nonin SenSmart 8204CA -oksimetrianturit molemmille otsalohkoille. Oikealla H500-mittalaitteen näyttöruutu. (www.nonin.com)

Kuvan 5 henkilöllä on oksimetrianturi molemmilla otsalohkoilla. BOPRA-tutkimuksessa dataa kerätään ainoastaan vasemmasta otsalohkosta. Mittalaite, josta kuvan 5 ruutukaappaus on otettu, on suunnilleen nykyaikaisen älypuhelimien kokoinen.

### 2.3 Autoregulaatiota kuvaavat indeksit

Autoregulaatiomekanismin toimintakykyä ja aivoverisuoniston reaktiivisuutta kuvaamaan on kehitetty monenlaisia fysiologisista mittauksista johdettuja laskennallisia tunnuslukuja. Aikatasoanalyysiin perustuvista indekseistä useimmat hyödyntävät perfuusiopainetta (CPP) ja verenvirtausta (CBF) kuvaavien mitta-arvojen välistä liukuvaa korrelaatiota. Tämän liukuvan korrelaation voidaan ajatella estimoivan autoregulaatiokäyrän (kuva 2) kulmakerrointa eri kohdissa. Kun korrelaatio on vahvasti positiivinen, on virtaus passiivisesti riippuvaista paineesta, eli autoregulaatiomekanismi ei pysty säätelemään virtausta. Vastaavasti matala korrelaatiokerroin ilmentää autoregulaatiomekanismin kykyä kompensoida painevaihtelun vaikutusta virtaukseen. [12.]

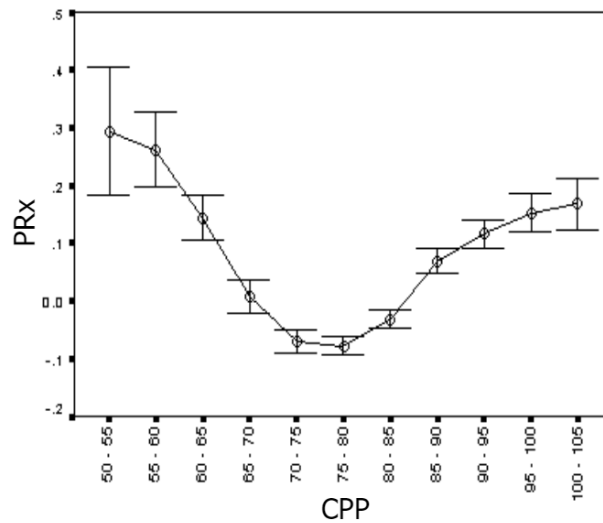
Taulukossa 1 on listattuna esimerkkejä tällaisista korrelaatioon perustuvista indekseistä. PRx (*Pressure Reactivity Index*) on valtimopaineen (MAP) ja kallonsisäisen paineen (ICP) välinen korrelaatio. Mx (*Mean Flow Index*) on perfuusiopaineen (CPP) ja Doppler-ultraäänimittauksella saadun keskimmäisen aivovaltimon virtausnopeuden keskiarvon (FVm) välinen korrelaatio. LDx vastaavasti on CPP:n ja laserdoppler-virtausmittauksen (LDF) välinen korrelaatio. [12.]

Taulukko 1: Korrelaatioon perustuvia autoregulaatioindeksejä.

Indeksi	Mittaustapa	Painearvo	Virtausarvo
PRx	Kallonsisäinen paine (ICP)	MAP	ICP
Mx	Ultraäänidoppler (TCD)	CPP	FVm
LDx	Laserdoppler (LDF)	CPP	LDF
COx	Kudosoksimetria (NIRS)	MAP	rSO <sub>2</sub>

Taajuus- ja siirtofunktioanalyysiin perustuvia autoregulaation mittareita ei käsitellä tässä opinnäytetyössä, koska BOPRA-tutkimuksessa tarkastellaan ainoastaan COx-indeksiä. Taulukossa 1 lueteltujen autoregulaatioindeksien tuntemus helpottaa tämän työn kannalta oleellisen tutkimustiedon ymmärtämistä.

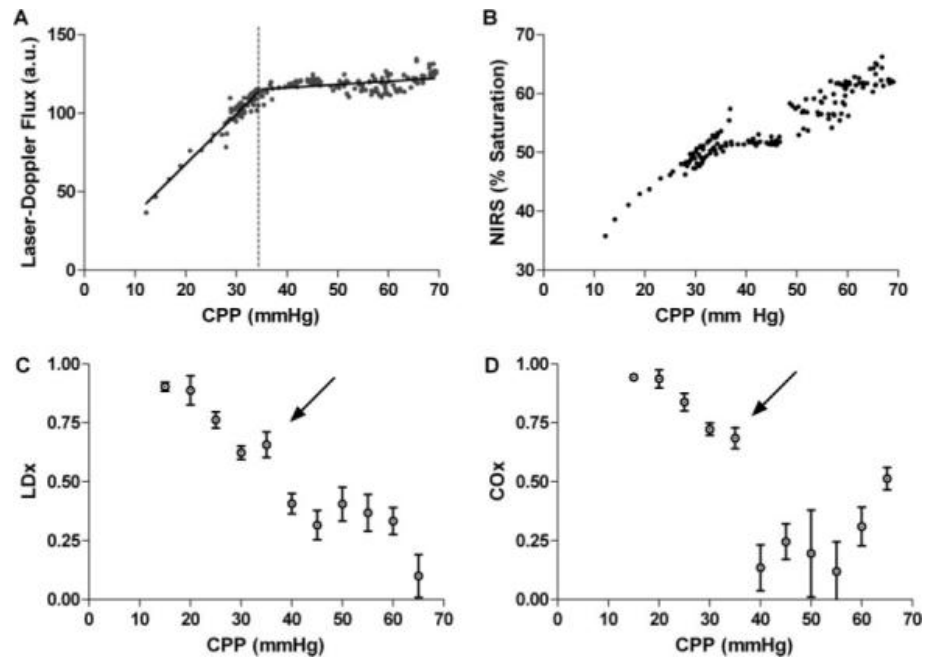
Piirrettäessä autoregulaatioindeksin kuvaaja verenpaineen funktiona, saadaan ihanteellisessa tilanteessa ylöspäin aukeava U-muotoinen käyrä, jonka minimikohdan voidaan ajatella olevan aivoverenkierrolle suotuisin verenpaine (kuva 6).



Kuva 6: PRx-indeksin graafinen esitys suhteessa perфуuסיopaineeseen (CPP). Minimikohta osoittaa tavoiteverenpaineen  $CPP_{opt}$ . [13].

Kuvassa 6 on Steiner ym. 2002 [13] -tutkimuksen kallovammapotilaista parhaiten toipuneen potilasryhmän PRx-indeksin keskiarvo eri verenpaineetasoilla. Alhaisilla ja korkeilla verenpaineetasoilla autoregulaatio häiriintyy ja PRx-indeksi kasvaa. Potilasryhmän ihanteellinen perфуuסיopaine vaikuttaa olevan hieman 80 elohopeamillimetrin alapuolella. Pisteyden yhteydessä olevat viiksimerkinnot kuvaavat PRx-indeksin arvojen hajontaa eri paineetasoilla.

Vallitsevan hoitosuosituksen mukaan aivovammapotilaan systolisen verenpaineen vähimmäistavoitearvo on aikuisilla 90 mmHg [14]. Hypotension on todettu olevan yhteydessä aivokudoksen vaurioitumiseen [2]. Yksilöity verenpaineen (MAP) hoitotavoite voi parantaa hoidon lopputulosta. Autoregulaatioindekseihin perustuvaa tavoiteverenpaineen yksilöintiä on tutkittu laajalti tehohoidon toimintaympäristössä [4; 8; 13]. Tässä opinnäytetyössä tarkasteltava COx-indeksi (*Cerebral Oximetry Index*) on valtimopaineen (MAP) ja aivokudoksen paikallisen happisaturaation ( $rSO_2$ ) välinen liukuva Pearson-korrelaatio (kuva 7).



Kuva 7: Porsaasta mitattujen COx- ja LDx-indeksien vertailu [15]. Kuviossa A on laser-Doppler-virtausmittauksen arvoja (LDF) kuvattuna verenpaineen suhteen. Kuviossa C on kuvion A LDF-arvojen perusteella laskettu LDx-indeksi. Kuviossa B vastaavasti NIRS-mittauksella saatu rSO<sub>2</sub>-arvo verenpaineen suhteen ja sen alapuolella kuviossa D COx-indeksi, joka on laskettu kuvion B lukemien perusteella.

Kuvan 7 ylemmissä kaavioissa on laboratoriokokeen optimoiduissa mittausolosuhteissa laser-Doppler- ja NIRS-mittausmenetelmillä tuotetut autoregulaatiokäyrät. Alemmissä kuvaajissa vastaavat LDx- ja COx-indeksit eri verenpainetasoilla. Verenpaineen ollessa matala, havaitaan korkeita indeksiarvoja. Brady ym. 2007 [15] -tutkimuksessa todetaan COx-indeksin pystyvän ilmaisemaan hypotensioon liittyvän autoregulaation menetyksen. Kuvassa 7 alaraja-arvo (LLA) on havaittavissa 35 mmHg perfuusiopaineen kohdalla. Selkeää autoregulaatioindeksien minimikohtaa ei näistä tuloksista pysty määrittämään. LDx-indeksi on odotetusti tarkkuudeltaan parempi. [15.]

### 3 Vaatimusmäärittely ja kehitysprosessi

Työssä alkuun pääsemiseksi oli sovittava tietosuojaan liittyvistä käytänteistä, selvitettävä syötedatan lukemiseksi tarvittavat toimenpiteet ja pyrittävä jo varhaisessa vaiheessa tunnistamaan mahdollisia riskejä.

NIRS-oksimetrin data on helposti luettavassa CSV-muodossa. Verenpainedata on ZOLL-monitoridefibrillaattorin tallentamassa JSON-muodossa, johon ei ole tarkkaa tietorakennekuvausta saatavilla. NIRS-dataan on lisäksi tarjolla kenttämuistiinpanojen perusteella tehtyjä korjauksia CSV-muodossa. Jatkokäsittelyä varten tuotettava data tallennetaan tilaajan toiveesta CSV-muotoon.

Potilasdatan aikaleimat tunnistettiin mahdolliseksi potilaan yksilöiväksi tiedoksi, joka tulisi eliminoida ennen tiedostojen käsittelyä tilaajaorganisaation ulkopuolella. Lisäksi huomioitiin riski potilastietojen tahattomasta päätyemisestä pilvipalveluihin tai versionhallintajärjestelmiin.

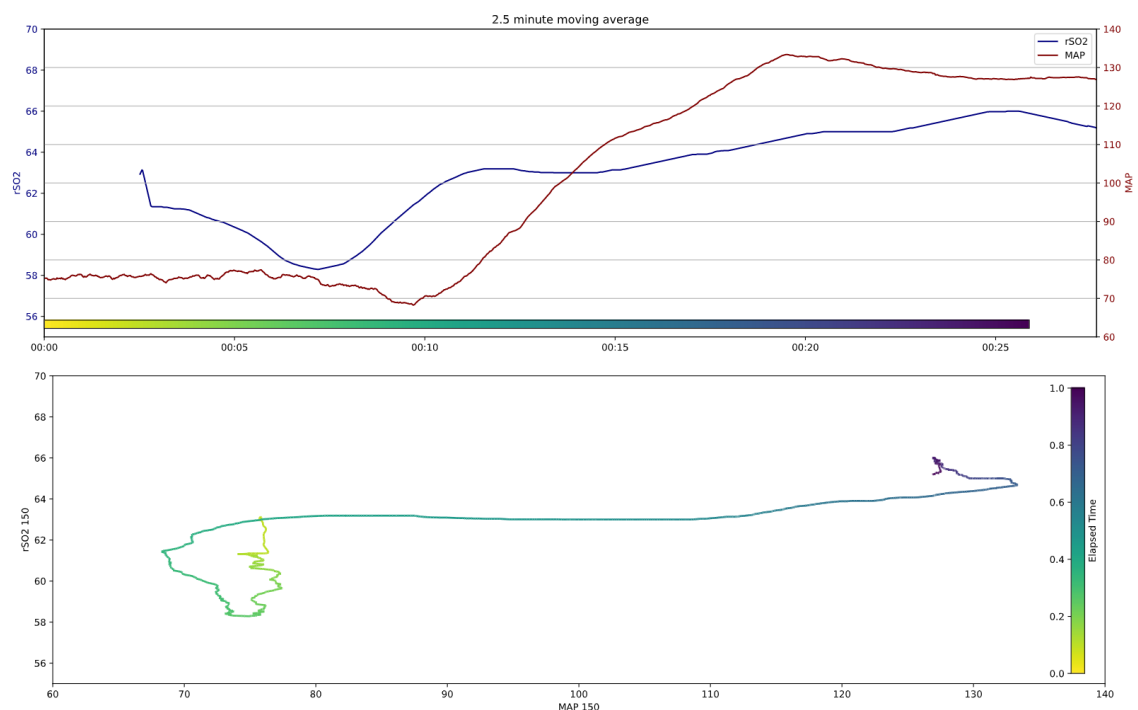
Data-analyysiprosessin suunnittelun lähtökohtana olivat tilaajan tarpeet. Pää tavoitteeksi valittiin COx-laskennan toteuttaminen. Työn kokeellisen luonteen takia myös vaatimusmäärittely miellettiin tietyiltä osin jatkuvaksi prosessiksi, sillä työn edetessä tieto karttuu ja tarpeet voivat muuttua.

Toteutettavassa data-analyysiympäristössä tietojen esikäsittelyn ja raporttien muodostamisen tulee olla mahdollisimman automatisoitua, jotta uuden datan lisääminen käy vaivatta [16]. Automatisoinnin ansiosta analyysiprosessi on myös läpinäkyvämpi ja helpommin toistettavissa [17]. Toteutuksessa pyritään noudattamaan tieteellisen tietojenkäsittelyn parhaita käytäntöjä [18].

Todennäköisenä ongelmalähteenä huomioitiin tiedonkeruuseen käytettyjen laitteiden kellonaikojen tuntematon epäsuhta, joka saattoi joissain tapauksissa olla useita minutteja. Haasteiksi arvellaan myös datapisteiden pientä määrää ensihoitotehtävien lyhyen keston vuoksi, sekä häiriöisyyttä, lähtöisin ensihoidon liikkuvasta toimintaympäristöstä.

Ensimmäisten koevisualisointien (kuva 8) tarkastelun jälkeen päätetään kokeilla erilaisien esikäsittelymenetelmien vaikutusta lopputulokseen tavoitteena selvittää, paljastuuko datasta autoregulaatiokäyrän muotoa. Kohdistusongelma pyritään korjaamaan joko automaattisesti ristikorrelaatiota hyödyntäen, tai visuaalisella tarkastelulla manuaalisesti. Verenpainedatassa esiintyvät häiriöt päädytään siivoamaan manuaalisesti.

Lisäksi sovittiin, että kokeillaan, tuottaako CO<sub>x</sub>-indeksin piirtäminen verenpaineen funktiona odotusten mukaisia U-muotoisia, nousevia tai laskevia käyriä. Tämä koe suoritetaan myös koko potilasaineistolla mukaellen Aries 2012 [4] -tutkimusta.



Kuva 8: Ensimmäisiä koevisualisointeja.

Kuvassa 8 on yritetty piirtää datasta autoregulaatiokäyrä. Ylemmässä kaaviossa on verenpaine- ja rSO<sub>2</sub>-signaalit ajan funktiona. Signaalien tasoituksessa on käytetty liiallisen pitkää 2,5 minuutin liukuvaa keskiarvoa. Alemmassa kaaviossa on MAP–rSO<sub>2</sub>-käyrä, jossa värikoodauksella on pyritty hahmottamaan kehitystä aika-akselilla. Havaittiin, ettei käyrämuotoinen esitys sovellu MAP–rSO<sub>2</sub>-havaintoparien tarkasteluun, koska syötearvojen vaihtelun myötä kaavion yli piirtyvä viivasyherö häiritsee tulkintaa. Jatkossa siirryttiin käyttämään hajontakuvioita MAP–rSO<sub>2</sub>-kaavioissa.

## 4 Laadulliset tavoitteet

Tietokoneohjelmat ovat nykyisin keskeisessä roolissa kaikilla tieteenaloilla. Verrattaen uutena teknologiana tietotekniikan oikeanlainen soveltaminen tutkimuksen välineenä ei ole itsestään selvä asia. Työn kokeellinen luonne tai osaamisen puute eivät käy syiksi laiminlyödä ohjelmistokehityksen peruskäytänteitä, kuten koodin ylläpidettävyyttä, versionhallintaa, virheenjäljitystä ja työnkulun automatisointia. Kuten laboratoriovälineet ja instrumentit, tulee tieteellisten johtopäätösten tuottamisessa käytetty koodi pystyä validoimaan. Puutteet tietojenkäsittelymenetelmien luotettavuudessa ovat johtaneet useisiin takaisinvetoihin ja korjauksiin korkean profiilin julkaisuissa, kuten *Science*, *PNAS*, *Journal of Molecular Biology*, *Ecology Letters*, *Journal of Mammalogy*, *Journal of the American College of Cardiology*, *Hypertension* ja *The American Economic Review*. [18; 19.] Tässä osassa esitetyt laatuominaisuudet ohjaavat työkalujen ja toteutustapojen valintaa.

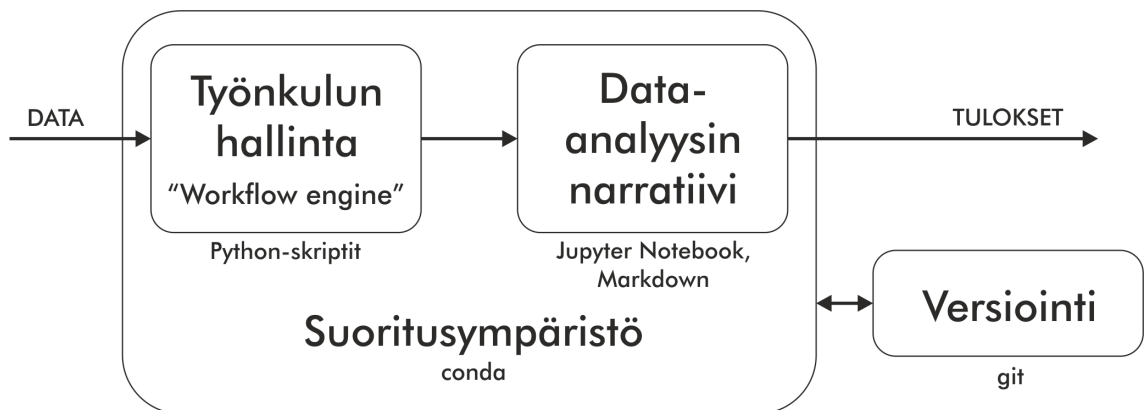
### 4.1 Validoitavuus ja läpinäkyvyys

Luotettavan data-analyysin ydin muodostuu tilastotieteen periaatteista. Hyviä tilastotieteen käytänteitä ovat tutkimuksen huolellinen suunnittelu ja läpivienti, graafiset ja numeeriset datan yhteenvedot, tutkittavan ilmiön ymmärrys, tulosten oikeanlainen tulkinta viitekehysessään, perusteellinen raportointi (liite 2) sekä kvantitatiivislooginen ymmärrys yhteenvedojen ja tunnuslukujen merkityksestä [20].

Edellä kuvailut tieteellisen tietojenkäsittelyn käytänteet mahdollistavat tilastotieteen ja lääketieteen asiantuntijoiden suorittaman validoinnin. Validointi muodostuu hankalaksi, mikäli ohjelmakoodi on dokumentoimaton, vaikealukuista, epäloogisesti etenevää tai rakenteeltaan hajanaista, saati jos analyysiprosessin osia on jätetty kokonaan raportoimatta [18].

## 4.2 Toistettavuus ja uusittavuus

Toistettavuus ja uusittavuus ovat tutkimustyön keskeisiä laatutekijöitä. Toistettavuus tarkoittaa kykyä toistaa koe siten, että samalla menetelmällä, samoissa olosuhteissa kerätyllä datalla saadut johtopäätökset ovat identtisiä. Data-analyysin toteuttava koodi tulee suunnitella siten, että se on kokonaisuudessaan helposti suoritettavissa alusta loppuun. Uusien analyysivaiheiden tai uuden datan lisäämisen tulee olla mahdollista ilman, että uudelleenajettavuus menetetään. Eksploratiiviseen data-analyysiin liittyy tarve kokeilla uutta sitä mukaa, kun tutkimusprosessin edetessä tehdään uusia huomioita. Tässä on tärkeässä roolissa koodin ylläpitokelpoinen ja skaalautuva rakenne (kuva 9) sekä versionhallinta. Ohjelmakoodin tulee toimia samalla tavoin riippumatta siitä, millä koneella se suoritetaan. Ohjelmakomponenttien ja suoritussympäristön eri versiot voivat aiheuttaa odottamattomia yllätyksiä. Näiden vakioinnissa voidaan hyödyntää suoritussympäristön hallintatyökaluja (esim. conda, virtualenv) ja virtuaaliympäristöjä (esim. Docker). [21; 22.]



Kuva 9: Ohjelmointityökalujen keskinäisiä suhteita data-analyysiprojektissa. Kuva mukaeltu [21].

Kuvassa 9 on havainnollistettu toistettavan data-analyysin mahdollistavan ohjelmointiympäristön rakennetta. Työnkulun hallintaa ohjaavat Python-skriptit suunnitellaan siten, että koko prosessi datasta tuloksiin voidaan suorittaa yhdellä komennolla. Kokeiluja voidaan suorittaa Jupyter notebook -tiedostoissa, kunnes niiden pohjalta löytyneet tuotantokelpoiset ratkaisut vakioidaan osaksi suoritussympäristöä.

Uusittavuudella tarkoitetaan sitä, että toinen tutkija voi saada samat tutkimustulokset samaa dataa käyttäen, kuvattua tutkimusprosessia noudattamalla. Uusittavuutta parantaa tutkimuksen läpinäkyvyys ja perusteellinen raportointi sekä datan saatavuus. Kliinises-

sä tutkimuksessa datan saatavuutta rajoittavat tietosuojakäytännöt ja salassapitovelvollisuus. [22.]

#### 4.3 Tietosuoja ja eettisyys

Kaikki tässä työssä tuotettu lähdekoodi on julkisesti tarkasteltavissa avoimen tutkimuksen periaatteen mukaisesti GitHub-palvelussa [23]. Potilastietoja käsitellään salassapitosäännösten ja -sopimuksen mukaisesti. Data sisältää tietueita, joiden perusteella potilaan voi mahdollisesti yksilöidä. Esimerkkejä tällaisista tiedoista ovat ikä, kellonajat ja päivämäärät. Yksilöivien tietojen poistamiseksi laadittiin skripti, joka nolaa eräät tietueet ja siirtää kaikki aikaleimat menneisyyteen siten, että alkuperäiset aikaleimat eivät ole niistä pääteltävissä.

## 5 Toteutus

Tässä osassa esitellään edellä kuvattujen vaatimusten ja laadullisten tavoitteiden saavuttamiseksi valitut työkalut ja toteutustavat sekä kuvaillaan joitain insinööriyön näkökulmasta kiinnostavia teknisen toteutuksen yksityiskohtia.

### 5.1 Työkalut

Toteutuksessa käytettäväksi ohjelmointikieleksi valittiin **Python**, joka on tekijälle ennestään tuttu. Python luotiin 1980-luvun lopulla skriptaus- ja opetuskäyttöön. Kieli on sittemmin noussut suureen suosioon ohjelmoijien ja tutkijoiden keskuudessa. Sille on kehitetty joukko data-analyysiin ja tutkimukseen tarkoitettuja komponentteja, jotka ovat laajalti liike-elämän ja tiedeyhteisön käytössä. [24.]

Python ei sellaisenaan sisällä numeeriseen laskentaan soveltuvia vektori- tai taulukko-rakenteita. **NumPy** on Python-moduuli, joka mahdollistaa matriisi- ja taulukkomuotoisen numeerisen datan tehokkaan käsittelyn ja sisältää useita matemaattisia funktioita. [24.]

**Pandas** on kaksiulotteisen taulukkodatan käsittelyyn tarkoitettu Python-moduuli. Keskeisin sen tarjoama työkalu on `DataFrame`-olio. [24.] Pandasin tarjoamia toimintoja ovat tiedosto- ja tietokantarajapinnat datan lukemiseen ja kirjoittamiseen, datan yhdistely, luokittelu, ryhmittely, uudelleenmuotoilu ja indeksointi sekä monipuoliset aikasarja-analyysin toiminnot, kuten liukuva aikaikkuna, taajuusmuunnos ja viivästys. Esimerkiksi CO<sub>x</sub>-indeksin laskennassa käytettävän liukuvan korrelaation voi pandasin avulla laskea helposti. Pandas hyödyntää toteutuksessaan NumPy-moduulia.

**Matplotlib** tarjoaa Python-ympäristöön työkalut kaaviokuvien piirtoon ja monipuoliseen datan visualisointiin [24]. **Seaborn** on Matplotlibille luotu rajapinta, joka helpottaa tiettyjen yleisten tilastollisten kaavioiden sekä värien ja tyylien määrittelyä.

**Jupyter notebook** on vuorovaikutteinen ohjelmointiympäristö. Siinä koodin osat kirjoite-

taan koodisoluihin, joita voi suorittaa yksitellen. Koodin tuottama tuloste tai kuvio tulee näkyviin solun alapuolelle. Koodisolujen lomaan voi lisätä kuvailevaa tekstiä, joten raportointi ja toimiva ohjelmakoodi voidaan laatia samaan tiedostoon. [24.] Jupyter notebookin rajoitteita ovat heikko yhteensopivuus versionhallinnan kanssa ja yksikkötestauksen hankaluus. Mahdollisuus suorittaa koodisoluja mielivaltaisessa järjestyksessä lisää ohjelmointivirheiden riskiä, sillä näin menettelemällä voi saada ohjelmointiympäristön ennalta arvaamattomaan tilaan. [25.]

**Papermill** on Jupyter notebook -tiedostojen parametroidun mahdollistava työkalu. Sen avulla Jupyter notebook -tiedoston sisältämän koodin voi suorittaa eri datasyötteillä. Tämä mahdollistaa useiden raporttien muodostamisen automatisoidusti käyttäen pohjana samaa Jupyter notebook -tiedostoa. [26.]

**Markdown** on merkintäkieli tekstitiedostojen muotoiluun. Markdown-formaattia noudattava sisältö on ihmiselle lukukelpoista sekä lähdemuotoisena raakatekstinä että typografisin muotoiluin esitettyinä. [27.] Jupyter notebookin tekstisolut tukevat Markdown-muotoilua.

**Conda** on ohjelmointiympäristön hallintaan ja ohjelmistopakettien asentamiseen tarkoitettu työkalu. Sen avulla voidaan määritellä tarkalleen käytössä olevat ohjelmistoversiot. Määritelmän mukainen ympäristö voidaan luoda helposti kaikille alustoille, joilla koodia suoritetaan. Tämä parantaa toistettavuutta ja karsii yhteensopivuusongelmia. [28.]

**Git** on Linus Torvaldsin kehittämä hajautettu versionhallintajärjestelmä. Versionhallintajärjestelmän tietovarastoon (*repository*) tallentuu koko projektin muokkaushistoria, ja se mahdollistaa mm. paluun aiempaan versioon sekä erillisten kehityshaarojen luonnin ja yhdistelyn. [29.] **GitHub**-palvelua voi käyttää git-tietovarastojen varmuuskopiointiin ja ohjelmakoodin julkaisemiseen [30].

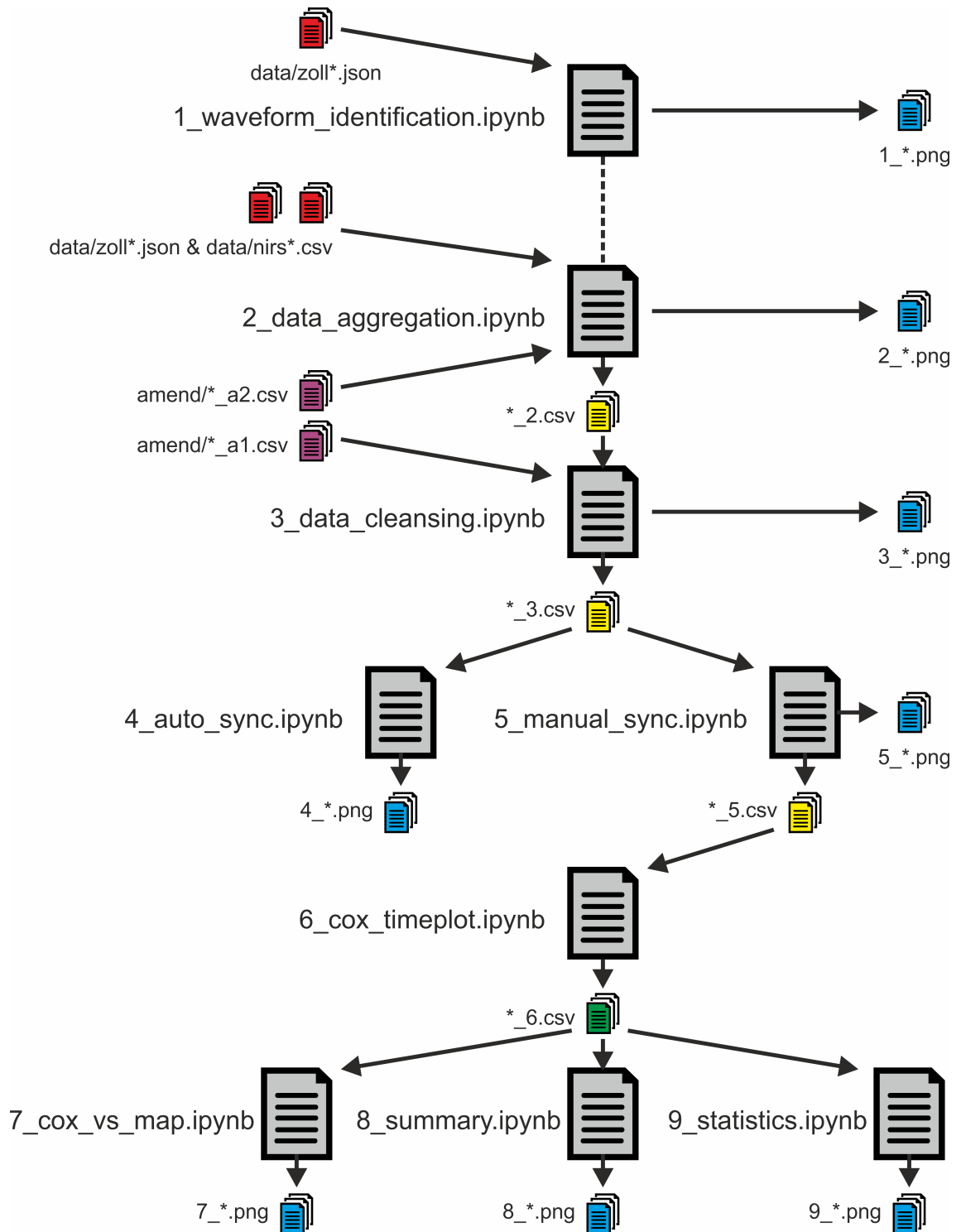
## 5.2 Suoritusympäristö

Datankäsittely raakadatan latauksesta loppuyhteenvetoon toteutettiin sarjana peräkkäisiä vaihteita. Kutakin vaihetta varten luotiin yksi Jupyter notebook -tiedosto suoritettavaksi jokaisen potilastapauksen syötedatalla Papermill-työkalua käyttäen. Tieto vaiheiden välillä siirtyy CSV-muotoisten välitallenteiden kautta. Tällainen ratkaisu katsottiin perustel-

luksi, sillä potilastapauksia on kymmeniä, ja ne kaikki ajetaan kahdeksan vaiheen läpi. Näin päästiin hyödyntämään sekä Jupyter notebookin joustavuutta analyysimenetelmien kokeilussa, että Python-skriptien tehokkuutta massa-ajojen suorittamisessa. Suunnittelun lähtökohdiksi valittiin ylläpidettävyys, luettavuus sekä mahdollisuus erilaisten menetelmien kokeilemiseen. Tärkeitä päämääriä olivat myös datan lisäämisen ja väliraportoinnin sujuvuus sekä kaikkien niiden vaiheiden automatisointi, joita ei ole välttämätöntä tehdä käsin. Suoritusympäristön vuokaavio on esitetty kuvassa 10. Kunkin vaiheen on tarkoitus käsittää yksi "hypoteesi–tulkinta–johtopäätös"-sykli. Näin menettelemällä ei yksittäinen notebook-tiedosto pääse kasvamaan hallitsemattoman laajaksi. [16.]

Valtaosa analyysin oikeellisuuteen vaikuttavasta koodista pyrittiin pitämään notebook-tiedostoissa, joista validoijan on sitä helppo tarkastella. Tiedostoformaattien käsittelyyn liittyvä koodi, sekä useammassa paikassa uudelleenkäytettävä koodi siirrettiin erillisiin Python-moduuleihin.

Hyvän tietojenkäsittelytavan mukaisesti mittalaitteiden tuottamaan dataan ei tehdä muutoksia käsin [18]. Manuaalisesti luodut merkinnät poistettavista datapisteistä ladataan erillisistä tiedostoista, ja niiden sisältö liitetään dataan ohjelmallisesti. Tämä helpottaa virheenjäljitystä.



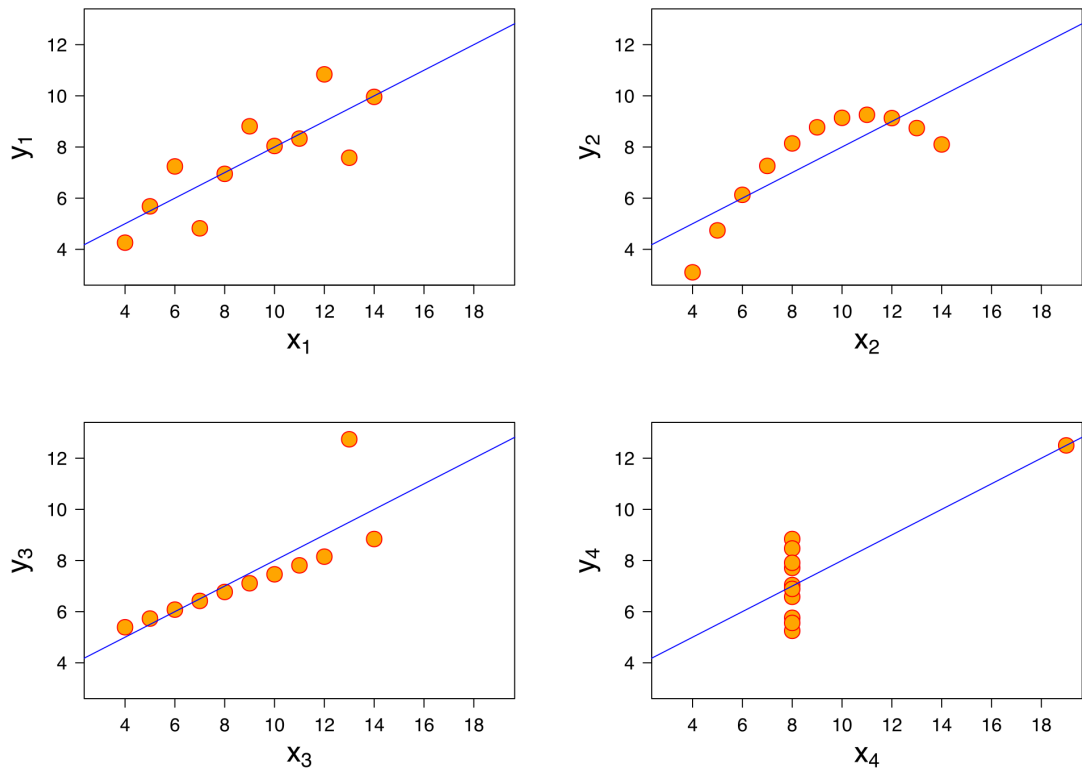
Kuva 10: Suoritusympäristön vuokaavio. Harmaalla värjätyt dokumentit ovat kunkin vaiheen pohjana toimivat Jupyter notebook -tiedostot. Syötteenä tuleva raakadata on merkitty punaisella. Dataan manuaalisesti tehdyt lisäykset on merkitty violetilla. Datavälitallenteet on merkitty keltaisella, ja lopullinen CSV-tallenne vihreällä. Raportointia varten tallennetut kuvatiedostot on merkitty sinisellä. Mallin mukautuvuudesta esimerkkinä toimii kohdan 4 (auto\_sync) tulokseton kokeilu, johon käytettyä koodia ei tarvitse poistaa, vaan se jätetään omaan haaraan osaksi projektia.

Eri vaiheiden muodostamat kuvatiedostot tallentuvat automaattisesti raportointihakemiston alle, josta Python-skripti kerää ne kunkin vaiheen koosteena toimivaan Markdown-tiedostoon. Hakemistoon luotiin yksinkertainen PHP-sivu, joka tarjoilee Markdown-tiedostot kuvineen verkkoselaimelle helposti katseltavassa muodossa. Tässä hyödynnettiin Parsedown-nimistä Markdown-jäsennintä. Näin raportointihakemiston sisällöstä saatiin itsenäisesti toimiva verkkosivu (liite 2), joka voitiin kopioida palvelimelle ja linkittää tutkijoille välitulosten tarkastelua varten. Lähdekoodin ja teknisen dokumentaation (liite 1) julkaisualustana käytettiin GitHub-palvelua.

### 5.3 Visualisointi

Eksploraatiivisessa data-analyysissä tutkitaan datan luonnetta usein visuaalisia menetelmiä käyttäen. Visualisoinnilla voidaan tiivistää laajat numeroaineistot ihmisen ymmärrettävään muotoon. Visualisointeja voidaan käyttää hypoteesien muodostamiseen aineistosta, josta niitä ei vielä ole. Vastaavasti visualisointeja voi hyödyntää varmistavaan analyysiin, jossa olemassa olevia hypoteeseja hyväksytään tai hylätään visuaalisen tarkastelun perusteella. [31.]

Graafinen tarkastelu voi paljastaa datasta ominaisuuksia, joita ei pelkkien tilastollisten tunnuslukujen perusteella ole mahdollista havaita. Mielenkiintoisena esimerkkinä tällaisesta tilanteesta toimii Anscomben kvartetti (kuva 11). Se muodostuu neljästä tietoaaineistosta, joilla on tunnuslukujen valossa keskenään samat ominaisuudet, kuten keskiarvo, korrelaatio, varianssi ja lineaarinen regressio. Aineistojen visuaalisesta esityksestä on kuitenkin helposti nähtävillä erilaiset muuttujien väliset yhteydet. [32.] Anscomben kvartetin muodostava data on esitetty taulukossa 2.



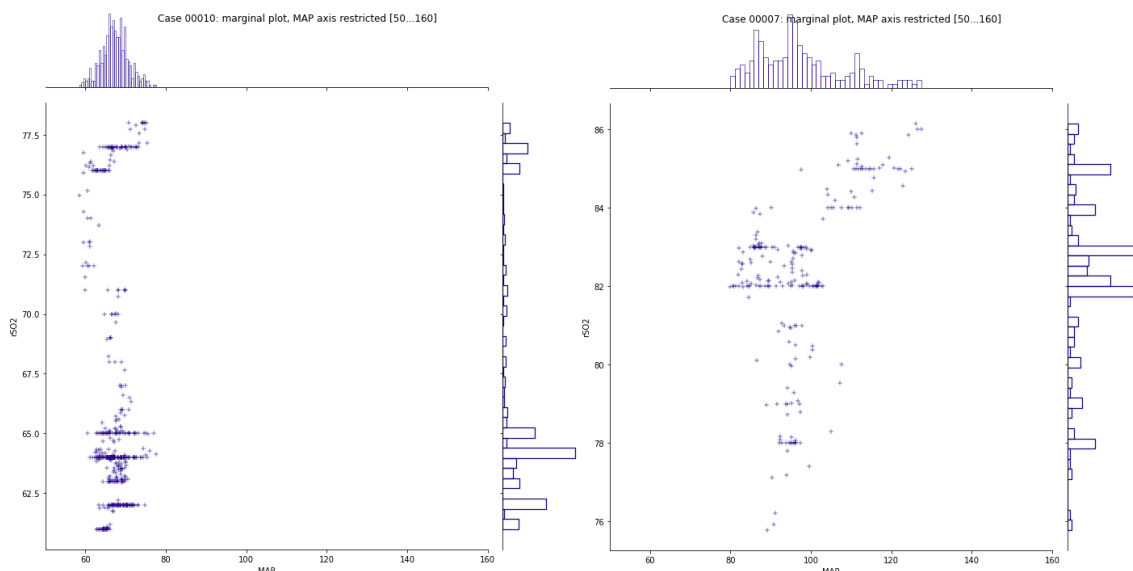
Kuva 11: Anscomben kvartetti. Kuviossa 1 nähdään lineaarinen yhteys, kuviossa 2 selkeä epälineaarinen yhteys, kuviossa 3 taas loivempi lineaarinen yhteys, jonka yksi poikkeava havainto saa matemaattisesti näyttämään samalta kuin kuvan 1 aineisto. Kuvion 4 muuttujien välillä taas ei ole lineaarista yhteyttä. Kaikilla aineistoilla on kuitenkin sama korrelaatiokerroin  $r_{xy} = 0,82$ .

Taulukko 2: Anscomben kvartetin data-aineisto

1		2		3		4	
$x$	$y$	$x$	$y$	$x$	$y$	$x$	$y$
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Kahden muuttujan välistä yhteyttä arvioitaessa on aina syytä tarkastella hajontakuviota (engl. scatter plot), jossa havaintoparit on kuvattu pisteinä [33]. Pisteparven muodosta voi-

daan tehdä päätelmiä muuttujien välisistä riippuvuuksista. Hajontakuvio helpottaa myös poikkeavien havaintojen tunnistamista. Kuvassa 12 on esitettyä hajontakuvioita kahdesta potilastapauksesta. Tässä työssä yhdistellään erilaisia visualisointitapoja kirjallisuutta ja tutkimustietoa hyödyntäen. Tavoitteena on hankkia mahdollisimman perusteellinen ymmärrys käsiteltävän datan luonteesta.



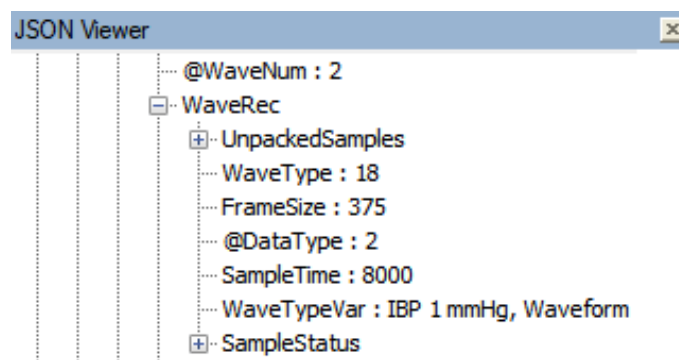
Kuva 12: Kahden potilastapauksen MAP–rSO<sub>2</sub>-hajontakuvioita. Vaaka-akseli kuvaa verenpainetta (MAP) ja pystyakseli aivokudoksen happisaturaatiota (rSO<sub>2</sub>). Pylväin esitetyt reunajakaumat havainnollistavat havaintopisteiden määriä akselin eri kohdissa. Vasemmanpuoleisen kuvion kaipa pisteparvi kertoo verenpainevaihtelun olevan riittämätöntä autoregulaatioilmiön havainnointiin. Oikeanpuoleisessa kuviossa taas nähdään havaintopareja laajemmalla verenpaineskaalalla sekä viitteitä autoregulaatiokäyrälle ominaisesta muodosta.

#### 5.4 Datan esikäsittely

Kuten data-analyysiprojekteissa on tyypillistä, tästäkin työstä suurin osa koostuu erilaisista datalle tehtävistä muunnoksista, jotka mahdollistavat tilastollisten yhteenvetojen laatimisen tulosten tulkintaa varten. Tässä osassa käydään läpi esikäsittelyn vaiheet, jotka muodostuivat hypoteesien sekä datasta tehtyjen havaintojen perusteella.

#### 5.4.1 Raakadatan jäsentäminen ja lataaminen

ZOLL-potilasmonitorin tuottamaan JSON-dataan ei ole saatavilla dokumentaatiota, joten tarvittavat tietueet oli etsittävä tiedoston rakennetta tutkimalla. Suuri, yksirivinen JSON-tiedosto muunnettiin luettavampaan muotoon lisäämällä rivitys ja sisennykset Pythonin `json.tool`-moduulin komentorivityökalulla. Tiedostot ovat kooltaan niin suuria, ettei niiden avaaminen onnistu useimmilla tekstieditoreilla. Vim-editori osoittautui toimivaksi miljoonia rivejä sisältävän tiedoston selailussa ja tekstihakujen tekemisessä. Kun oli näin saatu yleiskuva tiedoston rakenteesta, kirjoitettiin Python-skripti selvittämään tietueiden nimet ja lukumäärät. Tietueiden nimillä hakemalla paikannettiin haluttu sisältö. Datasta eristettiin yksittäisiä tietueita ja tarkasteltiin niiden rakennetta Notepad++ -editorin JSON Viewer -liitännäisen avulla (kuva 13).

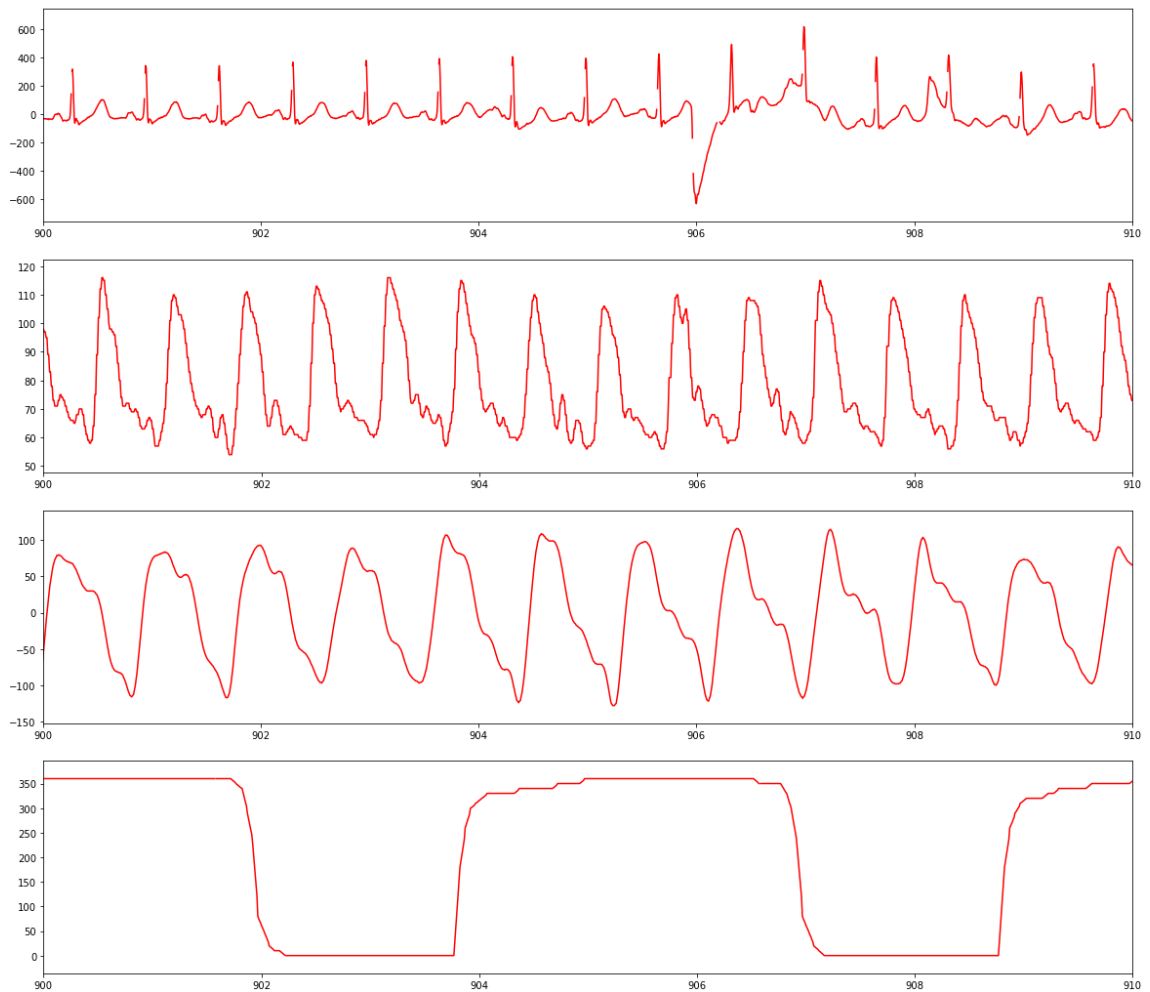


Kuva 13: JSON-dataa puunäkymässä.

Datan rakenne saatiin selville riittävällä tarkkuudella sisällön muuntamiseksi Pythonin data-analyysointityökalujen kanssa yhteensopivaan muotoon. NIRS-oksimetriadata on CSV-muodossa, joka on suoraan ladattavissa pandas-kirjastoa käyttäen.

#### 5.4.2 Aaltomuotojen tunnistaminen

ZOLL-defibrillaattorin tallentamat aaltomuodot tunnistettiin visuaalisesti tarkastelemalla, koska signaaleihin liitetyt kanavanumerot saattoivat vaihdella, eikä niiden perusteella voimut päätellä signaalin luonnetta. Aaltomuodoista piirrettiin jokaisen potilastapauksen osalta kuvan 14 mukaiset näytteet tunnistamista varten.



Kuva 14: ZOLL-defibrillaattorin tallentamia fysiologisia signaaleja.

Kuvassa 14 yllimpänä EKG-käyrä. Toisena ylhäältä invasiivisen verenpainemittauksen pulssikäyrä. Kolmantena ylhäältä pulssioksimetrin tuottama käyrä, ja alimpana hengitysilman hiilidioksidipitoisuus. Invasiivipaineen ja pulssioksimetrin käyrät ovat aaltomuodoltaan hyvin samankaltaiset. Ne voidaan kuitenkin erottaa toisistaan vaihteluvälin perusteella.

#### 5.4.3 Aikasarjojen tasoittaminen

Liukuva keskiarvo on aikasarja-analyysin perusmenetelmä, jolla voidaan tasoittaa aikasarjaa [34]. Se on äärellisen impulssivasteen alipäästösuodin, joka eliminoi aikasarjasta suuritaajuisia vaihtelukomponentteja. Liukuva keskiarvo voidaan esittää muodossa

$$y[n] = \frac{1}{L} \sum_{k=0}^{L-1} x[n-k] \quad (3)$$

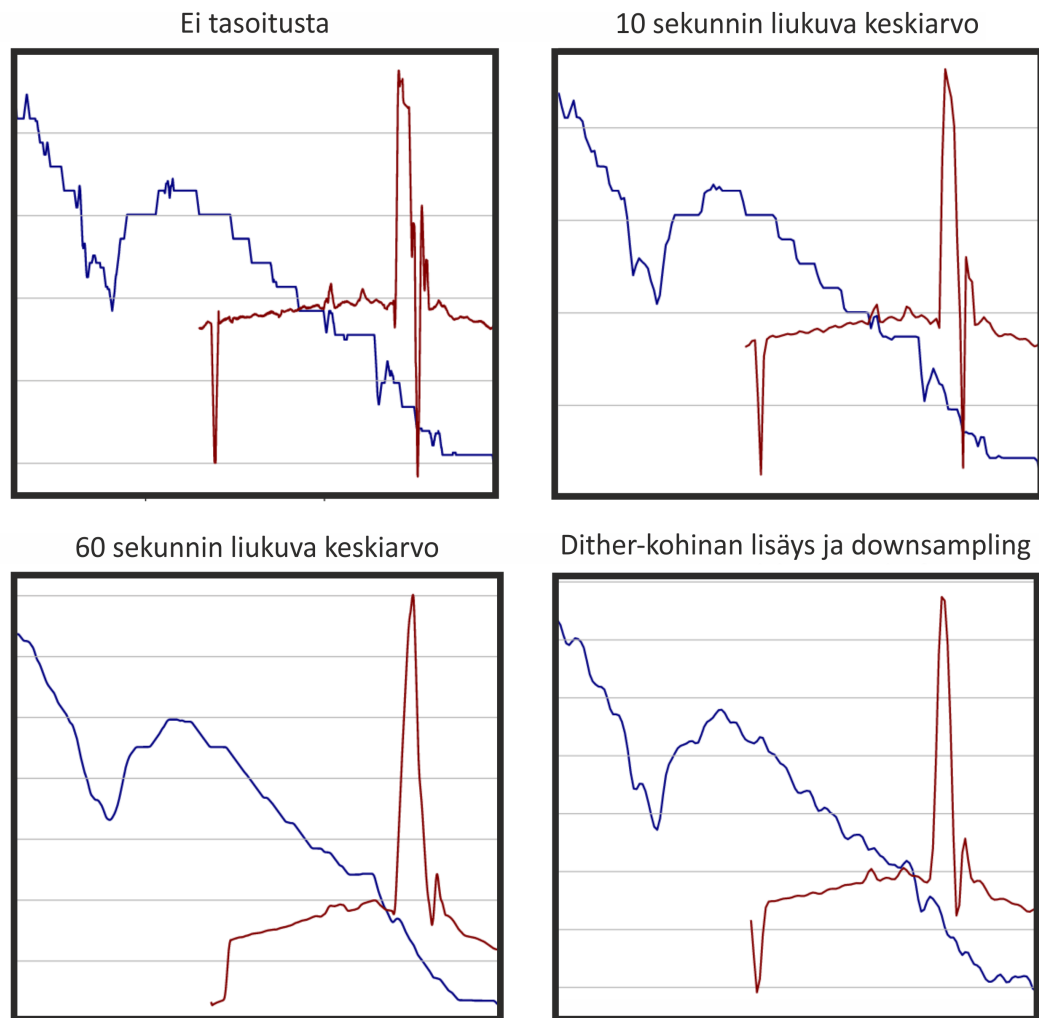
jossa  $L$  on aikaikkunan pituus,  $x$  on syötteeksi annettu aikasarja ja  $y$  lopputuloksena saatava keskiarvoistettu aikasarja [35]. Keskiarvo lasketaan  $L$  havaintopisteen välillä, joten aikasarjasta suodattuvat pois sellaiset vaihtelukomponentit, joiden kesto on aikaikkunaa lyhyempi [36]. Liukuva keskiarvo aiheuttaa signaaliin viiveen, joka on verrannollinen aikaikkunan pituuteen [35], joten yhdessä tutkittavat signaalit on syytä keskiarvoistaa samalla tavalla ajallisen kohdistuksen säilyttämiseksi. Pandas-kirjaston avulla aikasarjan liukuva keskiarvo saadaan seuraavasti:

```
df['x'].rolling(L).mean()
```

jossa `df` on pandasin `DataFrame`-olio. Operaatio suoritetaan sen sisältämälle `'x'`-nimiselle sarjalle.  $L$  on aikaikkunan pituus havaintopisteiden lukumääränä ilmaistuna.

Liukuvan keskiarvon laskennassa voidaan käyttää myös erilaisia *ikkunafunktioita*. Ikkunafunktio määrää painokertoimet aikaikkunan eri kohdille, sekä suodattimen impulssivasteen. Nelikulmioikkunasta puhuttaessa tarkoitetaan aikaikkunaa, jonka painokertoimet ovat kaikki yhtä suuria. Muita esimerkkejä ikkunafunktioista ovat kolmioikkuna sekä Hamming- ja Welch-ikkunat.

Määritettäessä keskivaltimopainetta (MAP) pulssikäyrästä ovat pulssin ja hengityksen verenpaineeseen aiheuttamat vaihtelukomponentit epätoivottuja [37]. Aikaikkunan pituudeksi valittiin 10 sekuntia, mikä riittää eliminoimaan matalienkin pulssi- ja hengitystaajuuksien aiheuttamat vaihtelut, mutta päästää läpi tutkimuksen kohteena olevat, matalataajuiset 0,005...0,05 Hz alueella olevat komponentit [15; 38]. Keskiarvoistus poistaa myös NIRS-mittalaitteen alhaisesta resoluutiosta johtuvaa rSO<sub>2</sub>-signaalin kvantisointisäröä (kuva 15).



Kuva 15: Kvantisointisärön suodatus eri parametrein.

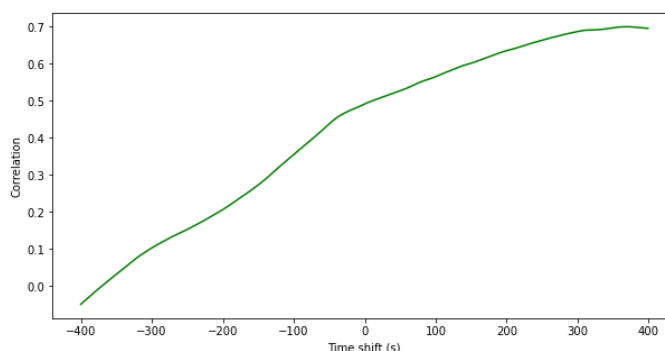
Tiedonkeruussa käytetyn  $rSO_2$ -arvoa mittaavan NIRS-laitteen resoluutio on 1 %. Joidenkin potilastapausten  $rSO_2$ -vaihteluväli on vain muutamia prosentteja. Tämä aiheuttaa signaaliin merkittävästi kvantisointisäröä joka näkyy signaalissa kulmikkaina muotoina. Kvantisointisärön aiheuttamat äkilliset muutokset signaalissa häiritsevät COx-indeksin määrittystä. Koska kyseessä on puhtaasti mittaustekninen häiriö, jolla ei ole yhteyttä mitattaviin fysiologisiin ilmiöihin, pyritään se suodattamaan mahdollisimman tehokkaasti pois. Kuvassa 15 on esitetty, miltä sinisen  $rSO_2$ -käyrän kvantisointisärö näyttää ilman signaalin tasoittamista, 10 ja 60 sekunnin liukuvalla keskiarvolla tasoitettuna sekä downsampling-tekniikalla käsiteltynä. Vaikuttaa siltä, että portaittaisen muodon pois suodattamiseksi vaaditaan 60 sekunnin aikaikkuna liukuvalla keskiarvolle. Pitkää aikaikkunaa käytettäessä

suodattuu myös tutkittavasta signaalista pois lyhyen aikavälin muutoksia. Ihanteellista olisi, jos kaikki 0...0,05 Hz taajuusalueella olevat muutokset saataisiin säilytettyä.

Koska tutkittavien signaalivaihtelujen ylärajataajuus  $f_s = \frac{1}{20} Hz$ , on vähimmäisnäytetiheys eli Nyquist-taajuus  $f_{nyquist} = 2f_s = \frac{1}{10} Hz$  [39]. NIRS-mittalaitteen näyteenottotaajuus on 1 Hz, joten signaali voidaan *desimoida* alemmalle näyteenottotaajuudelle. Kun signaaliin lisätään kohinaa (*dithering*), keskiarvoistetaan ja desimoidaan, paranee resoluutio ainakin teoriassa. Kuvan 15 oikeassa alakulmassa desimointikokeilu -1...1 kokonaislukukohinalla ja 50 sekunnin Hamming-ikkunan liukuvalla keskiarvolla.

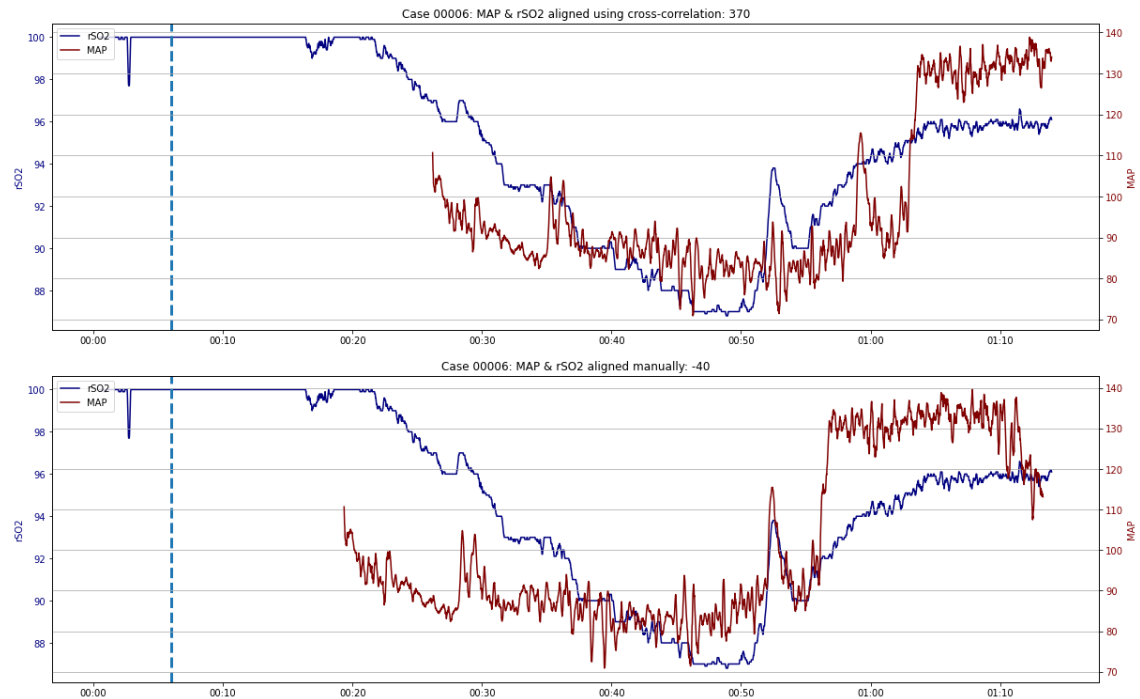
#### 5.4.4 Aikasarjojen kohdistaminen

MAP- ja rSO<sub>2</sub>-signaalien ajallisen kohdistusvirheen korjaamiseksi kokeiltiin ristikorrelaation maksimikohdan etsintää erisuuruisten viivästysten joukosta. Hyväksyttävän kohdistusvirheen ehdottoman maksimin katsottiin tutkimustiedon perusteella olevan noin kymmenen sekuntia, sillä tutkittavan ilmiön taustalla olevat verisuoniston reaktiiviset tapahtumat ovat kestoltaan 20–180 sekuntia [38]. Kuvassa 16 on erään tapauksen ristikorrelaation käyttäytyminen, kun verenpainesignaalia viivästetään -400...+400 sekuntia suhteessa rSO<sub>2</sub>-signaaliin.



Kuva 16: MAP–rSO<sub>2</sub>-ristikorrelaatio viivästyksen funktiona.

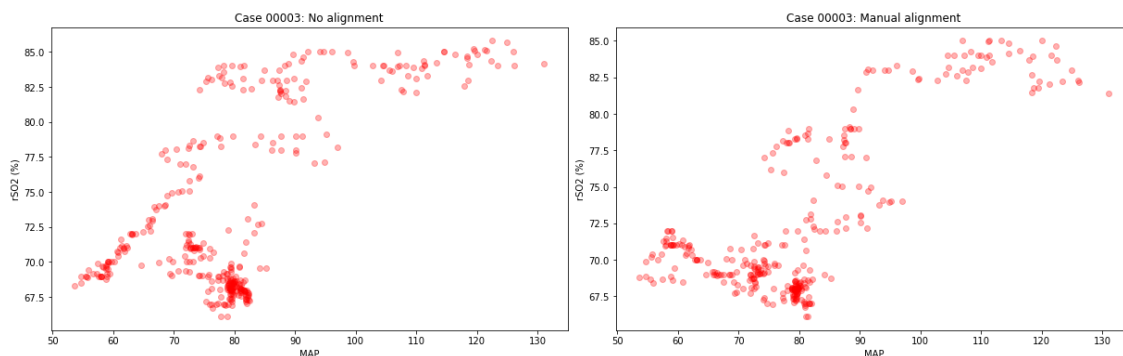
Maksimi löytyy 370 sekunnin viivästyksellä. Kuvassa 17 on ylempänä ristikorrelaatiomenetelmällä tehty, ja alempana manuaalisesti suoritettu kohdistus.



Kuva 17: Kohdistus ristikorrelaatiomenetelmällä ja manuaalisesti.

Esitettyssä tapauksessa ristikorrelaation tuottama kohdistus on selvästi virheellinen. Siniisen rSO<sub>2</sub>-käyrän voi olettaa reagoivan verenpaineen muutoksiin. Käyrien yhteisten muotojen perusteella signaalit on helppo kohdistaa manuaalisesti melko hyvällä tarkkuudella, vaikkakin silti kyseessä on tosiasiasa arvaus. Ristikorrelaatiomenetelmä ei siis sovellu näiden signaalien kohdistamiseen. Tämä on järkeenkäypää, sillä toimiva autoregulaatio nimenomaan vähentää signaalien välistä korrelaatiota.

Läheskään kaikissa tapauksissa ei manuaaliseen kohdistukseen löydy signaaleista yhtä selkeitä maamerkkejä kuin kuvan 17 esimerkissä. Tällöin on mahdotonta varmistua analyysin virheettömyydestä. Tällaisten signaalien välisistä suhteista ei sovi tehdä minkäänlaisia johtopäätöksiä. Näiden kokeilujen perusteella on suositeltavaa eliminoida kellovirhe jo datankeruuvaiheessa. Oikean kohdistuksen vaikutusta datan tulkintaan on havainnollistettu kuvassa 18.



Kuva 18: Kohdistuksen vaikutus MAP–rSO<sub>2</sub>-kuviioon. Vasemmalla puolella on erään potilastapauksen MAP–rSO<sub>2</sub>-hajontakuviio ilman syötesignaalien kohdistusta. Oikealla puolella on sama kuviio, mutta syötesignaalit kohdistettuina.

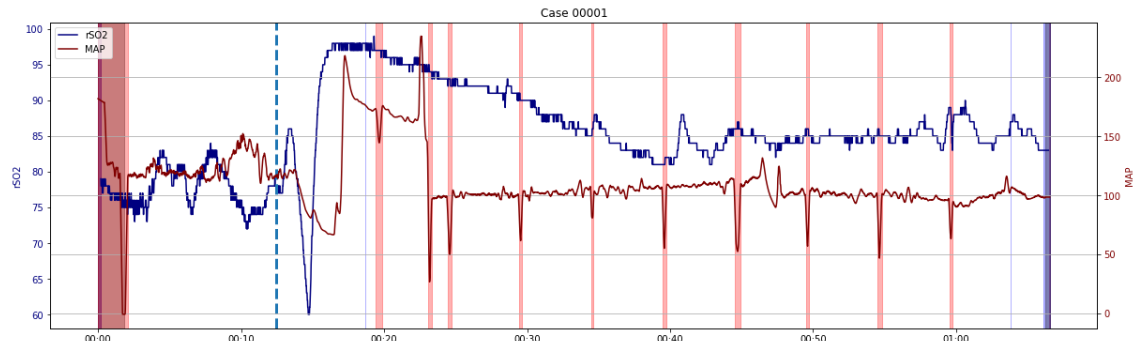
Kuvassa 18 on saman potilastapauksen MAP–rSO<sub>2</sub>-pisteparvi ilman syötesignaalien kohdistusta ja silmämääräisesti kohdistettuna. Kuvista voi yrittää tunnistaa kuvien 1 ja 2 mukaisia autoregulaatiokäyrän ominaisuuksia, kuten nousevia osuuksia sekä autoregulaatioalueen ylä- ja alarajaa. Signaalien kohdistuksella näyttää olevan vaikutusta tulkinnan lopputulokseen.

Pandasin `Series.shift()`-metodi hylkää olemassa olevan aikaindeksin ulkopuolelle siirtyvät arvot. Kohdistuksen toteutusta muokattiin siten, että näille arvoille tehtiin tilaa `DataFrame`:en laajentamalla aikaindeksiä. Näin saadaan mukaan kaikki informaatio mitalaitteiden kiinnitys- ja irrotushetkistä käyttöön johtolankoina selvittäessä oikeaa kohdistusta potilaskertomusten perusteella. Esikäsittelyvaiheet, mukaan lukien ajallinen kohdistus, tulee olla lyötynä lukkoon ennen lääketieteellisten johtopäätösten tekoon siirtymistä. Mikäli yksiselitteistä kohdistuspistettä ei löydy, voidaan potilastapaukseen liittyvä data joutua hylkäämään.

#### 5.4.5 Virheellisten arvojen poisto

Invasiivinen verenpainemittari on liikeherkkä. Tämän käsittelyvaiheen tarkoituksena on merkitä hylättäviksi selkeästi virheelliset lukemat. Näiden arvojen luotettava tunnistaminen sääntöperusteisesti tai algoritmisesti on haastavaa. Osa virhearvoista on helppo tunnistaa kliinisen kontekstin perusteella (kuva 19). Riittävän laadukkaaseen lopputulokseen ei kuitenkaan päästä ilman tarkempaa potilaskertomuksiin perehtymistä. Oleellista tä-

män vaiheen kannalta on toteutettu mekanismi, jolla arvoja voidaan merkitä poistettaviksi. Poistettavien datapisteiden yksityiskohtainen määrittely tehdään myöhemmin yhteistyössä tutkijoiden kanssa, kun potilasasiakirjat saadaan tarkasteltaviksi.



Kuva 19: Poistettaviksi merkattuja virheellisiä verenpainearvoja punaisella värillä korostettuna. Viiden minuutin välein toistuva verenpaineen mansettimittaus vaikuttaa rajoittavan verenvirtausta raajaan, josta invasiivipainetta mitataan.

## 5.5 COx-indeksin laskenta

Osassa 2 esiteltiin COx-autoregulaatioindeksin määrittely keskivaltimopaineen (MAP) ja aivokudoksen happikyllästeisyyden  $rSO_2$  välisenä liukuvana Pearson-korrelaationa. COx-indeksin laskennassa käytettävä Pearsonin korrelaatiokerroin on kahden muuttujan välisen lineaarisen riippuvuuden mitta. Se on muuttujien kovarianssin ja niiden keskihajontojen tulon osamäärä. [40].  $N$ -kokoisen otoksen Pearson-korrelaatio  $r_{xy}$  muuttujien  $x$  ja  $y$  välillä voidaan ilmaista seuraavasti:

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (4)$$

jossa  $\bar{x}$  ja  $\bar{y}$  ovat muuttujien otoskeskiarvot. Kovarianssin jakaminen keskihajontojen avulla normalisoi  $r$ :n vaihtelun välille  $[-1, 1]$ . Korrelaatiokertoimen itseisarvo ilmaisee lineaarisen riippuvuuden voimakkuutta, ja etumerkki kertoo sen suunnan. Arvo  $r = 1$  merkitsee täydellistä lineaarista riippuvuutta, kun taas  $r = -1$  merkitsee täydellistä negatiivista lineaarista riippuvuutta. [41.]

Kaavassa 3 esitettyä liukuvaa keskiarvoa vastaavasti voidaan korrelaatiokerroin määrittää liukuvaa aikaikkunaa käyttäen koko aikasarjalle (kaava 5):

$$r_{xy}[n] = \frac{\sum_{i=n}^{n+L-1} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=n}^{n+L-1} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=n}^{n+L-1} (y_i - \bar{y})^2}} \quad (5)$$

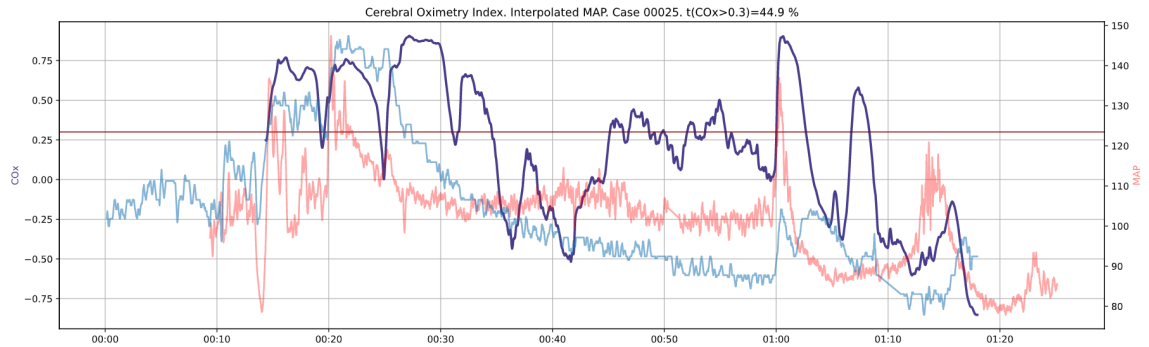
jossa  $L$  on aikaikkunan pituus ja  $n$  havaintoparin  $(x_n, y_n)$  järjestysluku. Korrelaatiokerroin siis lasketaan aikaikkunan sisältämälle otokselle. Aikaikkunaa siirretään eteenpäin kohtaan  $n + 1$  ja lasketaan uusi korrelaatiokerroin. Tätä toistetaan, kunnes aikasarjan loppu saavutetaan.

Pandas-kirjaston avulla kahden aikasarjan välinen liukuva korrelaatio saadaan laskettua seuraavasti:

```
df['x'].rolling(L).corr(df['y'])
```

jossa `df` on pandasin `DataFrame`-olio. Operaatio suoritetaan sen sisältämien `'x'`- ja `'y'`-nimisten sarjojen välillä.  $L$  on aikaikkunan pituus havaintopisteiden lukumääränä ilmaistuna.

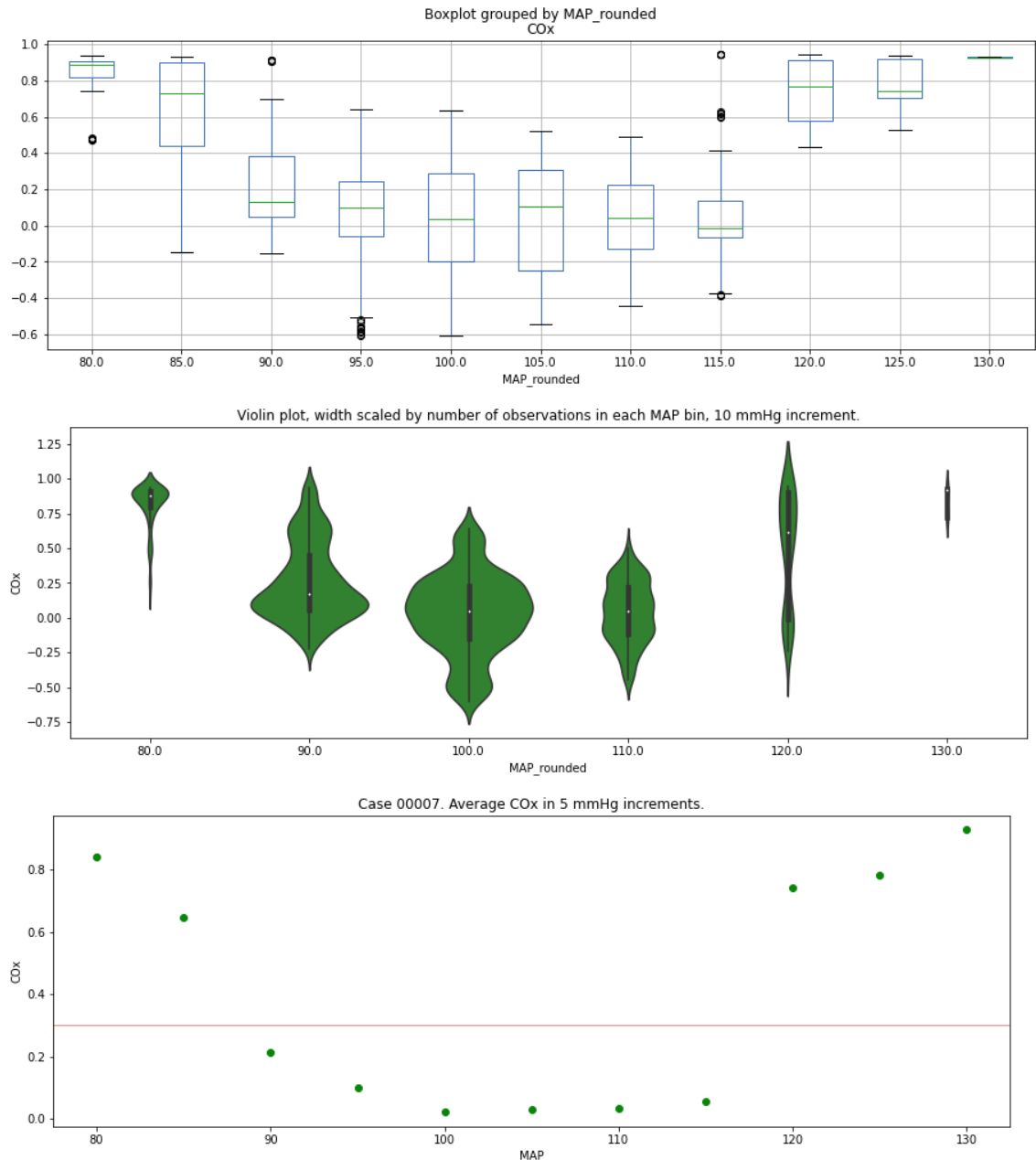
MAP–rSO<sub>2</sub>-korrelaation eli COx-indeksin voidaan mieltää estimoivan autoregulaatiokäyrän kulmakerrointa. Vahvasti positiivinen COx-indeksi tulkitaan viitteeksi autoregulaatiomekanismin kyvyttömyydestä kompensoida valtimopaineen muutosta. Kuvassa 20 on esitetty erään potilastapauksen COx-indeksin aikakuvaaja syötesignaaleineen.



Kuva 20: COx-indeksi laskettuna viiden minuutin aikaikkunaa käyttäen. Kuvasta voidaan havainnoida liukuvan korrelaation käyttäytymistä. Kun MAP ja  $rSO_2$  liikkuvat samaan suuntaan, COx-indeksi saa positiivisia arvoja. Vastaavasti, kun syötesignaalit liikkuvat eri suuntiin, COx-indeksi saa negatiivisia arvoja. Näyteväli on yksi sekunti, joten tässä liukuvan korrelaation aikaikkunassa sisältää 300 havaintoparia. Syötesignaalit on tasoitettu kymmenen sekunnin liukuvaa keskiarvoa käyttäen ja puuttuvat arvot paikattu lineaarisella interpoloinnilla.  $rSO_2$ -signaalissa on edelleen havaittavissa NIRS-mittalaitteen rajallisesta resoluutiosta johtuvaa kantikkuutta, joten signaalin keskiarvoistuksen aikaikkunaa voisi kokeilla pidentää.

## 5.6 Valtimopaineen optimointi

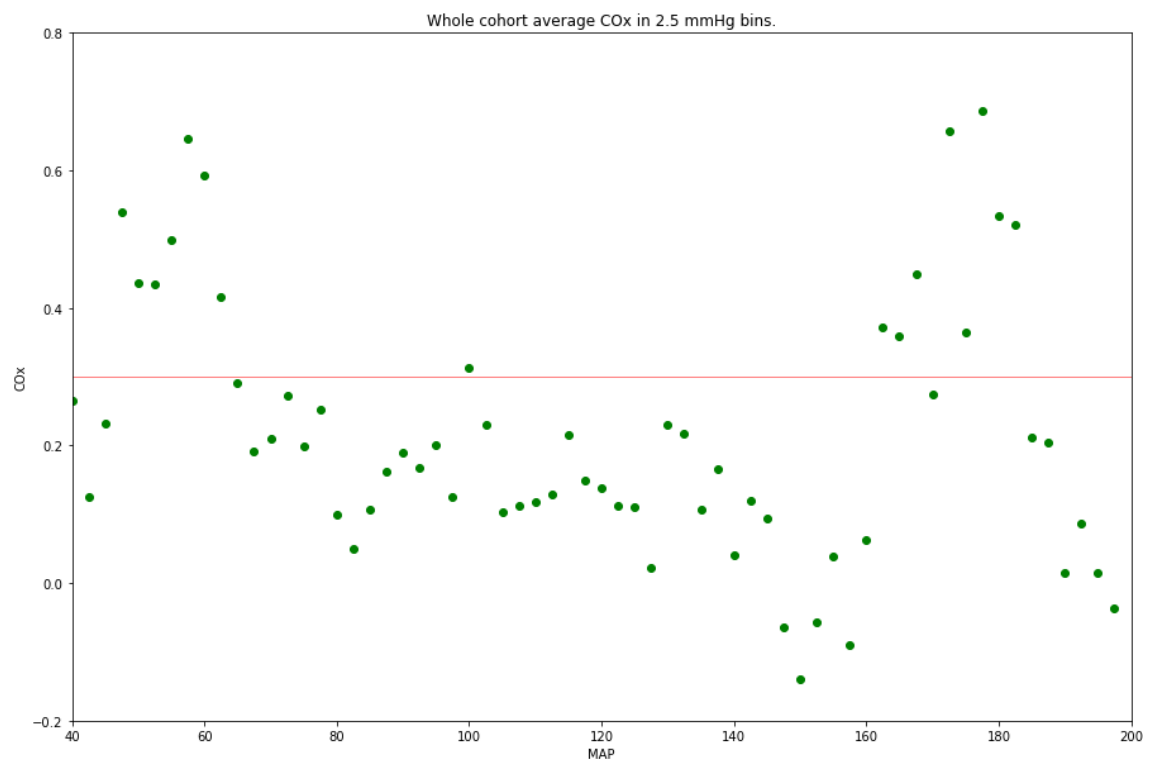
Kun COx-indeksin arvot kuvataan verenpaineen suhteen, oletetaan tuloksena olevan kuvan 6 mukainen U-muotoinen käyrä, jonka minimikohtaan ajatellaan olevan potilaan aivoverenkierron kannalta suotuisin verenpainetaso. Laskeva, nouseva tai U-muotoinen käyrä tukee käsitystä, että autoregulaatioilmiö on havaittavissa aineistosta. Jos tällainen muoto löydetään merkittävästä osasta aineistoja, saattaa minimikohtien tarjoamalla tiedolla olla lääketieteellistä arvoa. Kuvassa 21 on esitettyinä MAP–COx-kuvio kolmessa esitysmuodossa. Kyseessä on kuvan 12 oikeanpuoleisen hajontakuvion potilastapaus.



Kuva 21: Kolme esitystapaa MAP–COx-kuviolle. Ylimpänä olevassa laatikkokuviossa laatikoiden sisällä olevat vaakaviivat ovat mediaaneja. Laatikko muodostaa *kvartiilivälin*, eli puolet arvoista osuu laatikon sisäpuolelle. Laatikosta ylös ja alas osoittavat 'viikset' ovat matplotlibin oletusasetusten mukaisesti 1,5 kertaa kvartiilivälin pituiset. Niiden ulkopuolelle jäävät, ympyrällä kuvatut arvot katsotaan poikkeaviksi havainnoiksi (engl. outlier). Keskimmäisenä on ns. viulukuvio, joka havainnollistaa kunkin verenpainearvovälin havaintojen määrää ja sisäistä jakaumaa. Alimman kuvion pisteet esittävät COx-indeksin keskiarvoja eri verenpainetasoilla.

MAP–COx-kuvio muodostettiin myös koko aineistolle siten, että samassa pistejoukossa esiintyvät kaikkien potilastapausten kaikki havaintopisteet (kuva 22). Tässä odotuksena on havaintopisteiden suuren määrän ansiosta saada U-muotoinen käyrä huomatt-

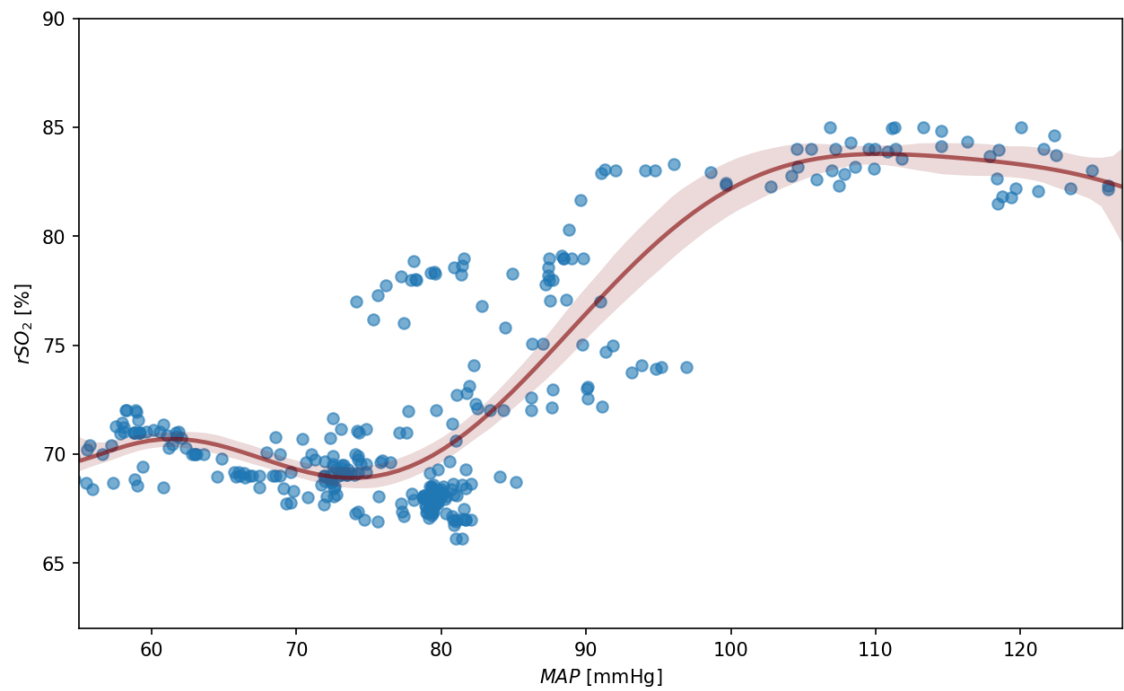
tavasti suuremmalla todennäköisyydellä kuin yksittäisestä potilastapauksesta. Menetelmä on keskeisen raja-arvolauseen nojalla pätevä, sillä syötteenä annettavat potilastapaukset ovat toisistaan riippumattomia [41, s. 303]. Tällaisia, lukuisia potilastapauksia käsittäviä kuviota voidaan hyödyntää esimerkiksi vertailtaessa eri potilasryhmien, kuten sydänpysähdys- tai kallovammapotilaiden autoregulaation käyttäytymistä.



Kuva 22: Koko aineiston kaikki tapaukset käsittävä MAP–COx-kuvio. Kuvaajasta on hahmotettavissa tietynlainen U-muoto. MAP:in kasvaessa 40:stä 140:een trendi vaikuttaa laskevalta, ja tämän jälkeen välillä 140...180 taas nousevalta. Yli 180 MAP arvoilla COx näyttäisi jälleen laskevan.

## 5.7 Viimeistelty visualisointi

Projektin aikana tilaajalla ilmeni tarve autoregulaatiokäyrää esittävälle kuvalle luentografiikkaa varten (kuva 23). Kuva haluttiin tuottaa BOPRA-aineistoa käyttäen, joten se luotiin tämän työn ohessa. Toiveena oli MAP–rSO<sub>2</sub>-hajontakuviota, johon tehdään käyräsovitusta. Käyräsovituksen ympärille haluttiin korostettuna alueena näkyviin 95 % luottamusväli.



Kuva 23: BOPRA-datasta poimittu tapaus, josta vaikuttaisi erottuvan autoregulaatiokäyrä

Kuvan 23 käyrä luotiin Seaborn-kirjaston `regplot()`-funktiota käyttäen. Vaikka tutkimusprosessin tämänhetkisessä vaiheessa ei vielä ole järkevää ryhtyä laajamittaisesti tuottamaan viimeistelyasteeltaan korkeatasoista grafiikkaa, voidaan todeta, että Pythonin työkaluilla sellainenkin on mahdollista.

## 6 Pohdinta

Tämän opinnäytetyön tuloksena syntyi käyttökelpoinen ohjelmakoodi BOPRA-tutkimukseen tarvittavan data-analyysin toteuttamiseksi. Työmenetelmien valinnat perusteltiin ammattialojen tietoperustaa ja kirjallisuutta hyödyntäen. Kehitysprosessissa ja sen dokumentoinnissa huomioitiin alusta lähtien keskeisinä tekijöinä tutkimustyön vaatimukset, kuten validoitavuus ja toistettavuus. Tutkimuseettisiä periaatteita ja tietosuojasäädöksiä noudatettiin koko projektin ajan. Kaikki työssä tuotettu materiaali on julkisesti saatavilla avoimen tieteen periaatteen mukaisesti. Uusittavuus pyrittiin takaamaan huolellisella dokumentoinnilla (liitteet 1 ja 2). Tuotosta sovelletaan meneillään olevassa tutkimustyössä, johon tämän opinnäytetyön tekijälle tarjoutui ainutlaatuinen tilaisuus osallistua. Saadut tulokset ovat olleet jossain määrin lupaavia, ja tutkimusta jatketaan edelleen opinnäytetyön valmistumisen jälkeen.

Projektin tulokset ovat olleet tilaajan edustajien mukaan innostavia. Kehitysprosessin dokumentointia luonnehdittiin laadukkaaksi, tekijän kykyä omaksua tieteellistä tietoa poikkeukselliseksi ja tutkimusyhteistyötä miellyttäväksi. Tätä kirjoitettaessa validointitapaamiset on vasta aloitettu, ja muutamia potilastapauksia päästy tutkimaan kliinisen kontekstin valossa. Tulokset ovat olleet lupaavia. Validointiprosessin myötä vahvistetaan alkuväestä datankäsittelyn lopulliset menetelmävalinnat, jonka jälkeen tuloksia voidaan alkaa hyödyntää tieteelliseen päättelyyn. Tekijä jatkaa työtä osana tutkimusryhmää.

Tuotosta validoidaan ja kaavioiden sisältöä tulkitaan parhaillaan potilastapausten kliinisen kontekstin valossa. Validoinnin yhteydessä havaittiin epäselvyys COx-indeksin laskennassa. Indeksien hetkellisarvoa vastaavaksi MAP-lukemaksi oli valittu aikaikkunan oikeanpuolimmainen arvo. Koska aikaikkuna sisältää useita verenpaine-arvoja, tulee täsmentää, mihin niistä kukin COx-indeksin arvo liitetään, vai käytetäänkö aikaikkunan keskiarvoa. Tämä asia varmistetaan laskentamallin lähteenä käytetyn tutkimuksen [4] tekijöiltä, jotta menetelmä tulee varmasti replikoitua oikein. Tutkimuksen tiedonkeruussa havaitun mittalaitteiden aikasynkronointiin liittyvän ongelman esiin tuominen johti korjaaviin toimenpiteisiin tutkimuksen meneillään olevassa tiedonkeruuprosessissa.

Ohjelmistokehityksen näkökulmasta tieteellisen työvälineen luomisessa tuntui haastavalta vaatimus koodin virheettömyydestä ja data-analyysin parametrien oikeellisuudesta ennen tulkintaan ja johtopäätösten tekoon siirtymistä. Iteratiivista ketterän kehityksen mallia on haastavaa soveltaa tämän tyyppiseen työhön. Kehittäjä on jossain määrin sidottu perinteisempään ns. vesiputousmalliin, jossa pääfokus on jatkuvan kehityksen sijasta lopullisen, toimivan tuotteen aikaansaamisessa. Eksploratiivisen data-analyysin tekeminen yhteistyössä kokousten aikana voisi sujuvoitua, jos käytössä olisi jokin interaktiivisen visualisoinnin työkalu, kuten Panel [42].

Tekijälle karttui runsaasti tietoa tutkittavasta fysiologisesta ilmiöstä, osaamista tieteellisen tietojenkäsittelyn menetelmistä, kokemusta data-analyysin käytännön toteutuksesta ja yhteistyöstä tutkijoiden kanssa. Työn edetessä hyödynnettiin runsaasti kirjallisuutta ja tieteellisiä artikkeleita. Erityisesti bioinformaattikoyhteisön julkaisuista löytyi runsaasti hyödyllistä tietoa dataintensiivisen tutkimuksen hallinnasta. Omaksuttavan tiedon määrä oli suuri, ja se vaati tiivistä paneutumista.

## Lähteet

- 1 BOPRA-tutkimuksen verkkosivu. 2020. Verkkodokumentti. FinnHEMS Ltd. <<https://www.boprastudy.fi>>. Luettu 30.10.2020.
- 2 Moerman A, De Hert S. Recent Advances in Cerebral Oximetry. Assessment of Cerebral Autoregulation with Near-Infrared Spectroscopy: Myth or Reality? F1000Research. 2017. <<https://doi.org/10.12688/f1000research.11351.1>>.
- 3 Nurmi J, Toivonen T, Laukkanen-Nevala P. BOPRA: Clinical Study Protocol, Version 2. FinnHEMS Ltd. 2019.
- 4 Aries MJH, Czosnyka M, Budohoski KP, Steiner LA, Lavinio A, Koliass AG, Hutchinson PJ, Brady KM, Menon DK, Pickard JD, Smielewski P. Continuous determination of optimal cerebral perfusion pressure in traumatic brain injury. Critical Care Medicine. 2012. <<https://doi.org/10.1097/ccm.0b013e3182514eb6>>.
- 5 Sand O, Sjaastad ØV, Haug E, Toverud KC, Bjålie JG. Ihminen: fysiologia ja anatomia. Helsinki: Sanoma Pro. 2014.
- 6 White H, Venkatesh B. Cerebral Perfusion Pressure in Neurotrauma: A Review. Anesthesia & Analgesia. 2008. <<https://doi.org/10.1213/ane.0b013e31817e7b1a>>.
- 7 Willie CK, Tzeng Y-C, Fisher JA, Ainslie PN. Integrative regulation of human brain blood flow. The Journal of Physiology. 2014. <<https://doi.org/10.1113/jphysiol.2013.268953>>.
- 8 CPP<sub>opt</sub>-tutkimuksen verkkosivu. 2020. Verkkodokumentti. CPP<sub>opt</sub> team. <<https://cppopt.org>>. Luettu 30.10.2020.
- 9 Stepan J, Hogue CW. Cerebral and Tissue Oximetry. Best Practice & Research: Clinical Anaesthesiology. 2014. <<http://www.ncbi.nlm.nih.gov/pmc/articles/pmc4258229/>>.
- 10 Watzman HM, Kurth CD, Montenegro LM, Rome J, Steven JM, Nicolson SC. Arterial and Venous Contributions to Near-infrared Cerebral Oximetry. Anesthesiology. 2000. <<https://doi.org/10.1097/00000542-200010000-00012>>.
- 11 Beerin ja Lambertin laki. 2021. Verkkodokumentti. Wikipedia. <[https://fi.wikipedia.org/wiki/Beerin\\_ja\\_Lambertin\\_laki](https://fi.wikipedia.org/wiki/Beerin_ja_Lambertin_laki)>. Luettu 31.1.2021.
- 12 Zeiler FA, Donnelly J, Cardim D, Menon DK, Smielewski P, Czosnyka M. ICP Versus Laser Doppler Cerebrovascular Reactivity Indices to Assess Brain Autoregulatory Capacity. Neurocritical Care. 2017. <<https://doi.org/10.1007/s12028-017-0472-x>>.
- 13 Steiner LA, Czosnyka M, Piechnik SK, Smielewski P, Chatfield D, Menon DK, Pickard JD. Continuous monitoring of cerebrovascular pressure reactivity

- allows determination of optimal cerebral perfusion pressure in patients with traumatic brain injury. *Critical Care Medicine*. 2002. <<https://pubmed.ncbi.nlm.nih.gov/11940737/>>.
- 14 Pakkanen T. The Impact of Emergency Medical Service Physicians on Patient Outcomes: With a focus on prehospital traumatic brain injury. Tampere University. 2020. <<http://urn.fi/URN:ISBN:978-952-03-1518-4>>.
- 15 Brady KM, Lee JK, Kibler KK, Smielewski P, Czosnyka M, Easley RB, Koehler RC, Shaffner DH. Continuous Time-Domain Analysis of Cerebrovascular Autoregulation Using Near-Infrared Spectroscopy. *Stroke*. 2007. <<https://doi.org/10.1161/STROKEAHA.107.485706>>.
- 16 Kazantsev V. 2017. Clean code in Jupyter notebooks, using Python. Verkkodokumentti. PyData. <<https://www.youtube.com/watch?v=2QLgf2YLLus>>. Luettu 30.10.2020.
- 17 Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, Wilm A, Holtgrewe M, Rahmann S, Nahnsen S, Köster J. Sustainable data analysis with Snakemake. Zenodo. 2020. <<https://doi.org/10.5281/zenodo.4067137>>.
- 18 Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK. Good enough practices in scientific computing. *PLoS Computational Biology*. 2017. <<https://doi.org/10.1371/journal.pcbi.1005510>>.
- 19 Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, Guy RT, Haddock D, Huff KD, Mitchell IM, Plumbley MD, Waugh B, White EP, Wilson P. Best Practices for Scientific Computing. *PLoS Biology*. 2014. <<https://doi.org/10.1371/journal.pbio.1001745>>.
- 20 Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016. <<https://doi.org/10.1080/00031305.2016.1154108>>.
- 21 Brown CT. 2018. Repeatability in Practice. Verkkodokumentti. <<https://hackmd.io/@ctb/BkreVNKYm>>. Luettu 1.1.2021.
- 22 Cobey K, Clement G. 2018. Data and reproducibility: The role of research institutions. Verkkodokumentti. <<https://publicationethics.org/news/data-and-reproducibility-role-research-institutions>>. Luettu 1.1.2021.
- 23 GitHub – BOPRA – Tämän opinnäytetyön lähdekoodi. 2021. Verkkodokumentti. Simo Mäkelä. <<https://github.com/makes/metropolia-bopra>>.
- 24 VanderPlas J. A Whirlwind Tour of Python. O'Reilly. 2016.
- 25 Pitfalls of Jupyter Notebooks. 2020. Verkkodokumentti. Aalto Scientific Computing. <<https://scicomp.aalto.fi/scicomp/jupyter-pitfalls/>>. Luettu 1.1.2021.
- 26 Papermill. 2021. Verkkodokumentti. nteract.io. <<https://papermill.readthedocs.io>>. Luettu 31.1.2021.

- 27 Markdown. 2021. Verkkodokumentti. Wikipedia.  
<<https://en.wikipedia.org/wiki/Markdown>>. Luettu 31.1.2021.
- 28 Conda – Package, dependency and environment management for any language. 2021. Verkkodokumentti. Continuum Analytics, Inc.  
<<https://conda.io>>. Luettu 31.1.2021.
- 29 Git. 2021. Verkkodokumentti. <<https://git-scm.com>>. Luettu 31.1.2021.
- 30 GitHub. 2021. Verkkodokumentti. <<https://github.com>>. Luettu 31.1.2021.
- 31 Luentomateriaali, Matti Gröhn – Tieteellinen visualisointi. 2003. Verkkodokumentti. Helsingin yliopisto.  
<<https://www.cs.helsinki.fi/u/htoivone/teaching/tutihak03/tietvis.pdf>>. Luettu 24.1.2021.
- 32 Tufte ER. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press. 2001.
- 33 Cohen PR. Empirical Methods for Artificial Intelligence. Cambridge, MA: MIT Press. 1997.
- 34 Käsitteet – Liukuva keskiarvo. 2021. Verkkodokumentti. Tilastokeskus.  
<[https://www.tilastokeskus.fi/meta/kas/liukuva\\_keskiar.html](https://www.tilastokeskus.fi/meta/kas/liukuva_keskiar.html)>. Luettu 18.1.2021.
- 35 Understanding Moving Average Filter with Python & Matlab. 2010. Verkkodokumentti. GaussianWaves.  
<<https://www.gaussianwaves.com/2010/11/moving-average-filter-ma-filter-2/>>. Luettu 18.1.2021.
- 36 Luentomateriaali, Ilkka Mellin – Aikasarjojen esikäsittely – Tasoitus: Liukuvien keskiarvojen käyttö. 2007. Verkkodokumentti. Aalto-yliopisto.  
<[http://salserver.org.aalto.fi/vanhat\\_sivut/Opinnot/Mat-2.3128/IMLuennot07/EAAS100.pdf](http://salserver.org.aalto.fi/vanhat_sivut/Opinnot/Mat-2.3128/IMLuennot07/EAAS100.pdf)>. Luettu 18.1.2021.
- 37 Aletti F, Hammond RL, Sala-Mercado JA, Chen X, O’Leary DS, Baselli G, Mukkamala R. Cardiac output is not a significant source of low frequency mean arterial pressure variability. *Physiological Measurement*. 2013.  
<<https://doi.org/10.1088/0967-3334/34/9/1207>>.
- 38 Steiner LA, Pfister D, Strebel SP, Radolovich D, Smielewski P, Czosnyka M. Near-Infrared Spectroscopy can Monitor Dynamic Cerebral Autoregulation in Adults. *Neurocritical Care*. 2009.  
<<https://doi.org/10.1007/s12028-008-9140-5>>.
- 39 Nyquistin teoreema. 2021. Verkkodokumentti. Wikipedia.  
<[https://fi.wikipedia.org/wiki/Nyquistin\\_teoreema](https://fi.wikipedia.org/wiki/Nyquistin_teoreema)>. Luettu 31.1.2021.
- 40 Manninen P, Ylén M. Tilastollisen päättelyn käytäntö – Tilastotiedettä soveltajille. Tampere: TI-tutkimuspalvelu. 2000.
- 41 Ross SM. *Introductory Statistics*. Cambridge: Elsevier. 2017.

- 42 Panel. 2021. Verkkodokumentti. HoloViz. <<https://panel.holoviz.org>>. Luettu 31.1.2021.

# Liite 1: Tekninen dokumentaatio

## Cerebral oximetry data analysis pipeline

Analysis and reporting framework to study the relationship between cerebral tissue oximetry (rSO<sub>2</sub>) and mean arterial pressure (MAP).

1. Requirements
2. Setup
3. Usage
4. Project Files
5. Output CSV Format

## Requirements

Miniconda or anaconda installation.

## Setup

Clone the repository to your computer:

```
git clone https://github.com/makes/bopra.git
cd bopra
```

To install the required software packages, run the following command that creates a conda environment.

```
conda env create --name bopra-cox --file environment.yml
```

or, to store the environment under project folder, you can use

```
conda env create --prefix ./env --file environment.yml
```

Activate the installed environment using

```
conda activate bopra-cox
```

or, if you installed the environment under project folder,

```
conda activate ./env
```

## Usage

Copy the raw data files into the `data` folder. NIRS files must be named `nirs[case id].csv`. ZOLL files must be named `zoll[case id].json`.

The `amend` folder must contain the files `[case id]_a1.csv`, `[case id]_a2.csv` and `timeshift.csv`.  
\* `_a1.csv` are comma delimited files containing column `Bad_MAP_manual` to exclude invalid MAP readings.  
\* `_a2.csv` are semicolon separated files containing column `HuonoSignaali2` to exclude invalid NIRS readings.

```
├─data
│   ├── nirs00001.csv
│   ├── zoll00001.json
│   ├── nirs00002.csv
│   ├── zoll00002.json
│   └── ...
├─amend
│   ├── timeshift.csv
│   ├── 00001_a1.csv
│   ├── 00001_a2.csv
│   ├── 00002_a1.csv
│   ├── 00002_a2.csv
│   └── ...
```

In `config.ini`, specify which `CaseIds` to process in each step. This allows for exclusion of datasets.

To execute all steps for all the cases, use the script `runall.ps1`. To do partial runs, use the `process.py` script. Image assets are stored in the `reports/images` folder. The `report_visuals.py` script gathers the image assets to markdown files for easy viewing. The `reports` folder is a self contained website for easy publishing.

## Project files

```
├─*.ipynb
├─anonymize.py
├─config.ini
├─environment.yml
├─process.py
├─README.md
├─report_visuals.py
├─runall.ps1
├─adhoc
├─amend
├─data
├─doc
├─output
├─reports
│   ├──images
│   ├──markdown
│   └─notebooks
└─utils
```

`environment.yml` specifies the `conda` environment, and what software packages to install. See section `setup` for instructions.

`runall.ps1`: A script to execute the entire pipeline from start to finish. Can be run in Anaconda Powershell prompt.

`*.ipynb`: The Jupyter notebooks for each analysis step. Used as papermill templates.

`process.py`: A script to process an analysis step. Usage: `python process.py <step>` for all cases defined in `config.ini`. To select a single case for processing, use the `--case <id>` argument. Uses papermill to do the work.

`report_visuals.py` outputs markdown to display the generated image assets.

`anonymize.py`: A script to obfuscate timestamps in data to prevent identification.

`data`: A folder containing input data - NIRS CSV and ZOLL JSON.

`amend`: A folder containing manually crafted additions to the data, i.e. markings for artefact removal and synchronization.

`output`: Output CSV, intermediate and final.

`reports`: Output directory for image assets and reports. This directory can be hosted on a PHP enabled web server for easy sharing. To view locally, PHP's built in web server can be used: `php -S localhost:8000`.

`utils`: Python modules containing helper functions for data loading and processing.

`adhoc`: Test scripts and other throwaway code.

`doc`: Additional documentation (Excel workbooks).

## Output CSV format

Missing values are denoted by `'`. Time resolution is 1 second.

- `Time`: Timestamp in format H:M:S, starting at 00:00:00
- `MAP`: Mean arterial pressure computed from pulse waveform using a 10 second moving average
- `Bad_MAP_auto`: `1` indicated blood pressure values marked invalid by the monitoring device, `0` otherwise
- `Bad_MAP_manual`: `1` indicated blood pressure values manually marked invalid, `0` otherwise.
- `rSO2`: Cerebral tissue oxygen saturation percentage measured using NIRS. Resolution 1 %.
- `Mark`: Value `1` indicates start of anaesthesia, `0` otherwise
- `Bad_rSO2_auto`: `1` indicated rSO<sub>2</sub> values marked invalid by the monitoring device, `0` otherwise
- `Bad_rSO2_manual`: `1` indicated rSO<sub>2</sub> values manually marked invalid, `0` otherwise
- `COx`: "Cerebral oximetry index" calculated using a 300 second rolling correlation between MAP and rSO<sub>2</sub>, with bad input values dropped out and imputed by linear interpolation.
- `MAP_delay`: The first value of this column indicates the MAP delay that was used to synchronize the input signals.

# Lite 2: Data analysis dokumentaatio

## Assessing autoregulation using cerebral oximetry in a prehospital setting

This project is an attempt to examine cerebral autoregulation using data recorded in a prehospital environment.

The two physiological signals used in this analysis are invasive blood pressure recorded using a ZOLL monitor-defibrillator, and left frontal lobe rSO<sub>2</sub> recorded each second using a Nonin H500 NIRS monitoring device. Data collection was done as a part of the [BOPRA study](#) by Finnish HEMS physicians during patient transport under anaesthesia.

rSO<sub>2</sub>-MAP plots will be rendered to see if a pattern of the Lassen autoregulation curve emerges in measurement data recorded in a prehospital environment. The value of COx index will be plotted over time, and the time spent above the threshold COx > 0.3 will be noted to assess association with mortality, morbidity and quality of life. Also, COx-MAP plots will be rendered to see if a U-shaped curve emerges, revealing a supposed individualized optimal MAP for cerebral perfusion. With sufficient pressure variation, it should be possible to detect LLA/ULA at some degree of reliability.

Expected challenges are the relatively short duration of patient contact resulting in a scarce number of data points, as well as the mobile recording environment with plenty of interference.

The author of this analysis has no knowledge of the underlying pathologies or other patient metadata. This information exists in the BOPRA database and can be associated with each patient case for further research.

The purpose of this document is to facilitate the assessment of methodological validity. For interpretation, please see [section 8](#) for key visualizations of each case and [section 9](#) for numerical statistics and whole-cohort charts. To assess the technical validity of this analysis, please refer to the source code at <https://github.com/makes/metropolia-bopra>.

1. [Waveform Identification](#)
2. [Data Aggregation](#)
3. [Data Cleansing](#)
4. [Automatic Time Synchronization](#)
5. [Manual Time Synchronization](#)
6. [COx Index Over Time](#)
7. [COx vs MAP](#)
8. [Summary](#)
9. [Statistics](#)

### 1: Waveform Identification

The ZOLL monitor-defibrillator stores the physiological signals recorded under six channel ids (0 to 5). The purpose of this step is to identify the nature of each signal to ensure correct processing.

Two visualizations were created for each channel. The first one shows the entire time span of the signal, and the second one shows a zoomed-in detail of ten seconds at the fifteen minute mark.

[View the images used for signal identification](#)

#### Hypothesis:

Each signal can be assumed to occupy the same channel id predictably across all cases. The invasive arterial pressure waveform required for the analysis can be found in all twelve cases.

#### Interpretation:

Signals identified for each case:

Case ID	Ch 0	Ch 1	Ch 2	Ch 3	Ch 4	Ch 5
1	BP	-	-	-	-	BP
2	BP	-	-	-	-	BP
3	EKG	BP	CO2	Pleth	-	-
4	EKG	CO2	Pleth	-	-	BP+CO2
5	EKG	BP	CO2	Pleth	-	-
6	-	BP	-	-	-	BP
7	-	BP	-	-	-	BP
8	EKG	BP	CO2	Pleth	-	-
9	-	BP	-	-	-	BP+CO2
10	EKG	BP	CO2	Pleth	-	-
11	EKG	CO2	Pleth	-	-	-
12	EKG	CO2	Pleth	-	-	No BP
13	-	BP	-	-	-	No BP
14	EKG	CO2	Pleth	-	-	No BP
15	-	BP	-	-	-	BP
16	-	BP	-	-	-	BP
17	EKG	BP	CO2	Pleth	-	-
18	EKG	BP	CO2	Pleth	-	-
19	-	BP	-	-	-	BP+CO2
20	-	BP	-	-	-	BP+CO2
21	-	BP	-	-	-	BP
22	-	-	-	-	-	-
23	EKG	CO2	Pleth	-	-	No BP
24	-	BP	-	-	-	BP
25	EKG	BP	CO2	Pleth	-	-
26	EKG	BP	CO2	Pleth	-	-
27	EKG	CO2	Pleth	-	-	No BP
28	EKG	CO2	Pleth	-	-	No BP
29	EKG	BP	CO2	-	-	BP+CO2
30	-	-	-	-	-	-
31	-	-	-	-	-	-
32	EKG	CO2	Pleth	-	-	No BP
33	EKG	CO2	Pleth	-	-	No BP
34	EKG	CO2	Pleth	-	-	No BP
35	EKG	CO2	Pleth	-	-	No BP
36	EKG	CO2	Pleth	-	-	No BP
37	EKG	CO2	Pleth	-	-	No BP
38	EKG	CO2	Pleth	-	-	No BP
39	EKG	CO2	Pleth	-	-	No BP
40	EKG	CO2	Pleth	-	-	No BP
41	-	BP	-	-	-	No BP
42	-	BP	-	-	-	BP
43	-	-	-	-	-	-
44	-	-	-	-	-	-
45	EKG	BP	CO2	Pleth	-	-
46	EKG	BP	CO2	Pleth	-	-
47	EKG	BP	CO2	Pleth	-	-
48	EKG	CO2	Pleth	-	-	No BP
49	-	BP	-	-	-	BP
50	EKG	CO2	Pleth	-	-	No BP
51	EKG	BP	CO2	Pleth	-	-
52	EKG	-	CO2	Pleth	-	BP+CO2
53	-	-	-	-	-	-
54	EKG	BP	CO2	Pleth	-	-
55	EKG	BP	CO2	Pleth	-	-
56	-	-	-	-	-	-
57	EKG	CO2	Pleth	-	-	No BP
58	EKG	BP	CO2	Pleth	-	-

#### Conclusion:

Cases 4, 11, 12, 14, 23, 27, 28, 32, 33, 34, 35, 36, 37, 38, 39, 40, 48, 50, 52 and 57 did not contain the required IAP waveform, and thus will be excluded from further analysis.

Case 31 produces an unresolved loading error and will be excluded for the time being.

Cases 22, 30, 43, 44, 53 and 56 fail this step due to a software bug, but contain IAP data.

Of the 58 datasets provided, 37 have IAP data available under channel 1. Sample rate of the IAP pulse waveform is 125 Hz.

### 2: Data Aggregation

In this step, mean arterial pressure is computed from the pulse waveform using a moving average of 10 seconds, then resampled to match the 1 Hz sample rate of the rSO<sub>2</sub> signal.

A csv file containing both signals will be produced for each case, as well as a visualization of each signal on a common time axis.

The csv file will also contain columns indicating bad sample quality reported by the device, and a `Mark` field indicating start of anaesthesia.

#### Hypothesis:

Data from the two separate source devices can be converted to a common sample rate of 1 Hz and aggregated into a single csv file.

#### Data:

[View the visualizations](#)

#### Conclusion:

CSV files useful for further analysis were produced.

### 3: Data Cleansing

In this step, anomalies and artefacts in MAP data will be identified by visual inspection of the signal. Any obviously invalid samples will be marked for exclusion.

Invalid values in rSO<sub>2</sub> data have already been marked by the researchers based on field notes and visual inspection. These markings will be incorporated from a separate file. Corrected `Mark` timestamps indicating start of anaesthesia will also be included from this file. The corrections are based on field notes.

A csv file containing the data from step 2, amended with markings described above, will be generated.

#### MAP datapoint exclusion criteria:

Any MAP values not already marked as invalid by the monitoring device, where one or more of the following apply:

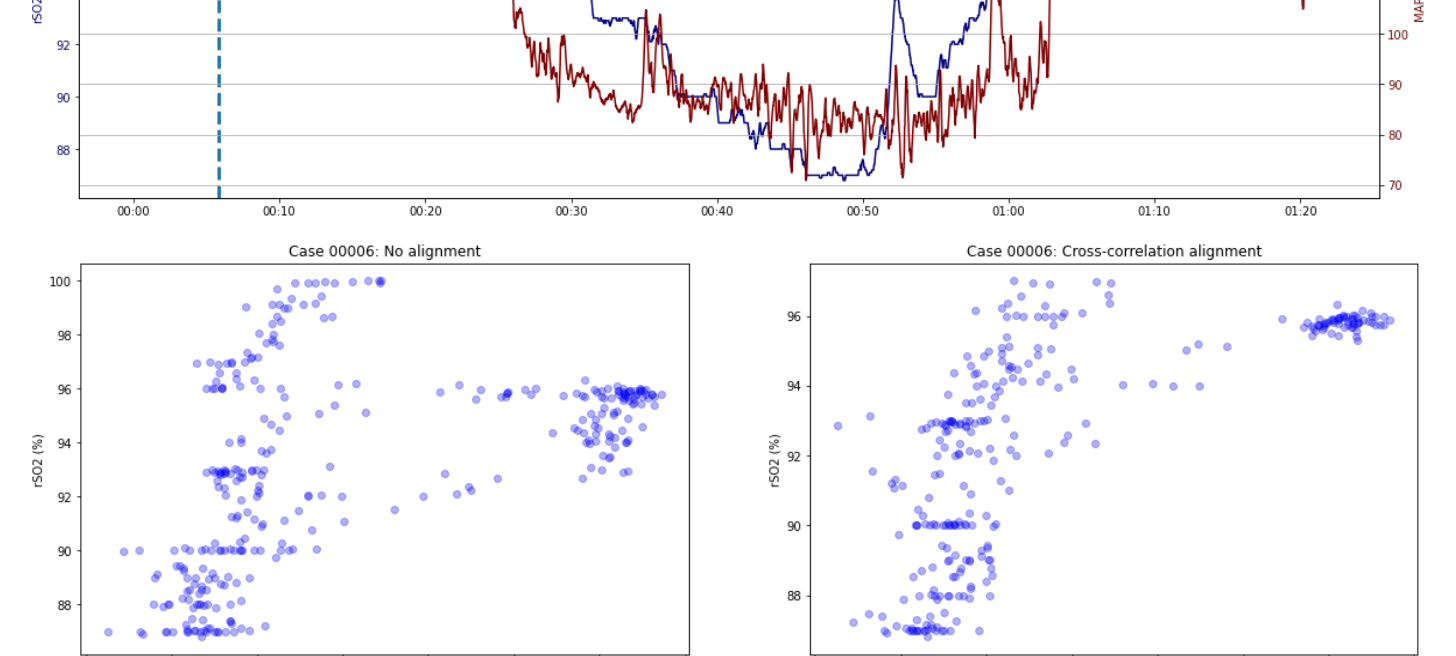
- Fluctuations at the start of IAP monitoring
- Sudden MAP surge above 180 mmHg AND immediate return to baseline recorded before the spike
- Sudden MAP drop below 40 mmHg, as would happen during arterial line flushing
- Periodically occurring MAP drop due to NIBP measurement

#### Hypothesis:

There will be artefacts in the blood pressure signal caused by movement and nursing procedures.

#### Interpretation:

- Blood pressure data manually marked for exclusion is shaded red.
- Blood pressure data automatically marked as low quality is shaded darker red/grey
- rSO<sub>2</sub> data manually marked for exclusion is shaded purple.
- rSO<sub>2</sub> data automatically marked as low quality is shaded blue.



[View all images](#)

#### Conclusion:

Case 24 was excluded from further analysis, as the rSO<sub>2</sub> signal was missing most of the time. Case 47 was excluded from further analysis, as no IAP data was recorded.

- Any abnormally high or abnormally low MAP readings were marked for exclusion.
- Periodically occurring drops in blood pressure were identified as a NIBP cuff restricting blood flow to the limb used for invasive blood pressure measurement. These values were marked for exclusion.
- The exclusion criteria are pending domain expert review.

A csv file containing the data described above was generated.

### 4: Automatic Synchronization

It has been noted that the two monitoring devices used for data collection may have a clock discrepancy of several minutes. This is identified as a potential source of error, as the vasogenic phenomena under analysis cycle within 20 to 180 seconds ([Steiner et al., 2008](#)).

#### Hypothesis:

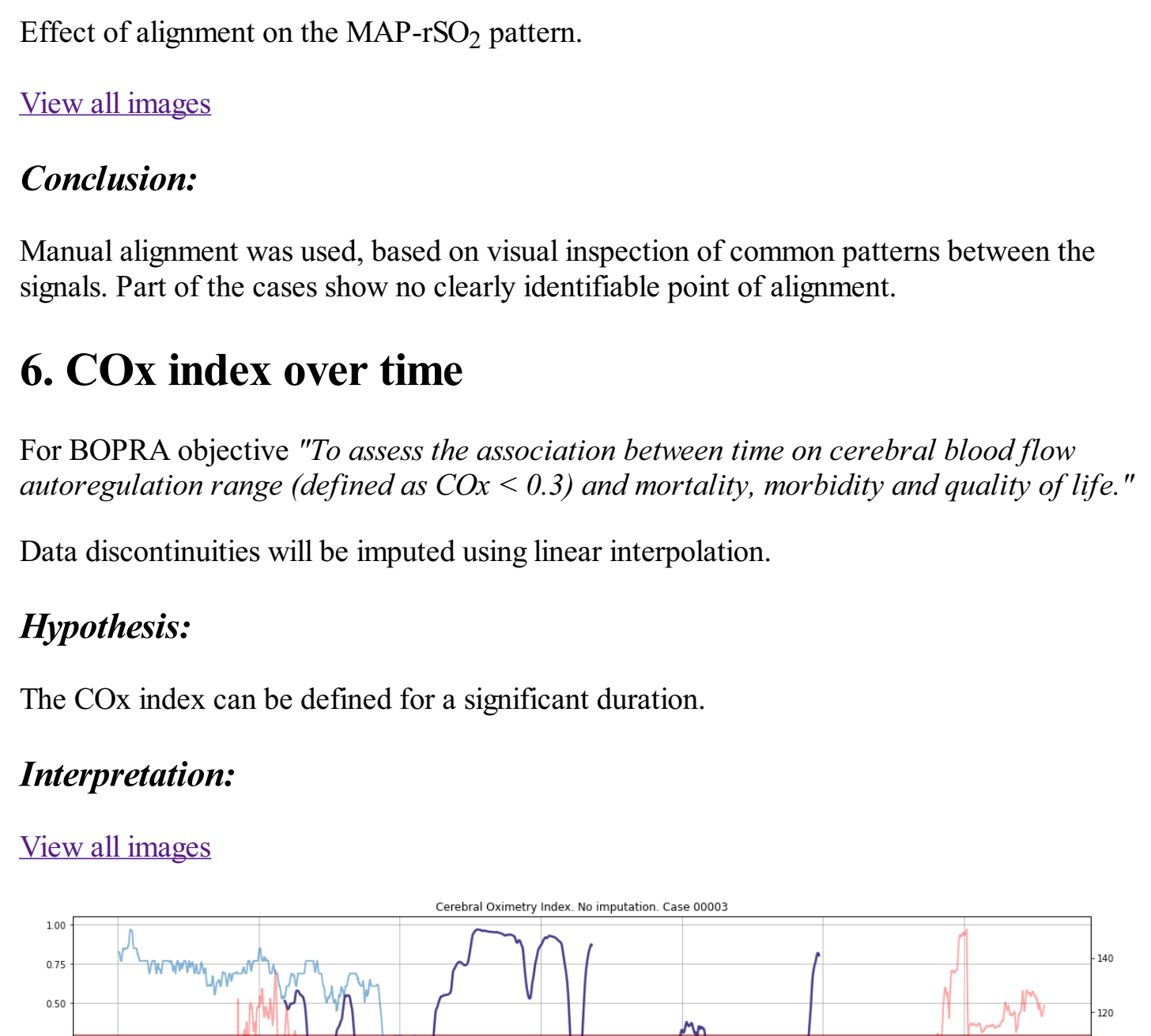
Automatic synchronization using cross-correlation can be used to reliably align the signals.

**Caveat:** Lack of correlation between the signals indicates good autoregulation.

#### Interpretation:

Here is an example auto-align attempt:

[View all images](#)



Effect of alignment on the MAP-rSO<sub>2</sub> pattern.

[View all images](#)

#### Conclusion:

Cross correlation produced sensible results in only part of the cases. The method is not suitable for this analysis.

### 5: Manual Synchronization

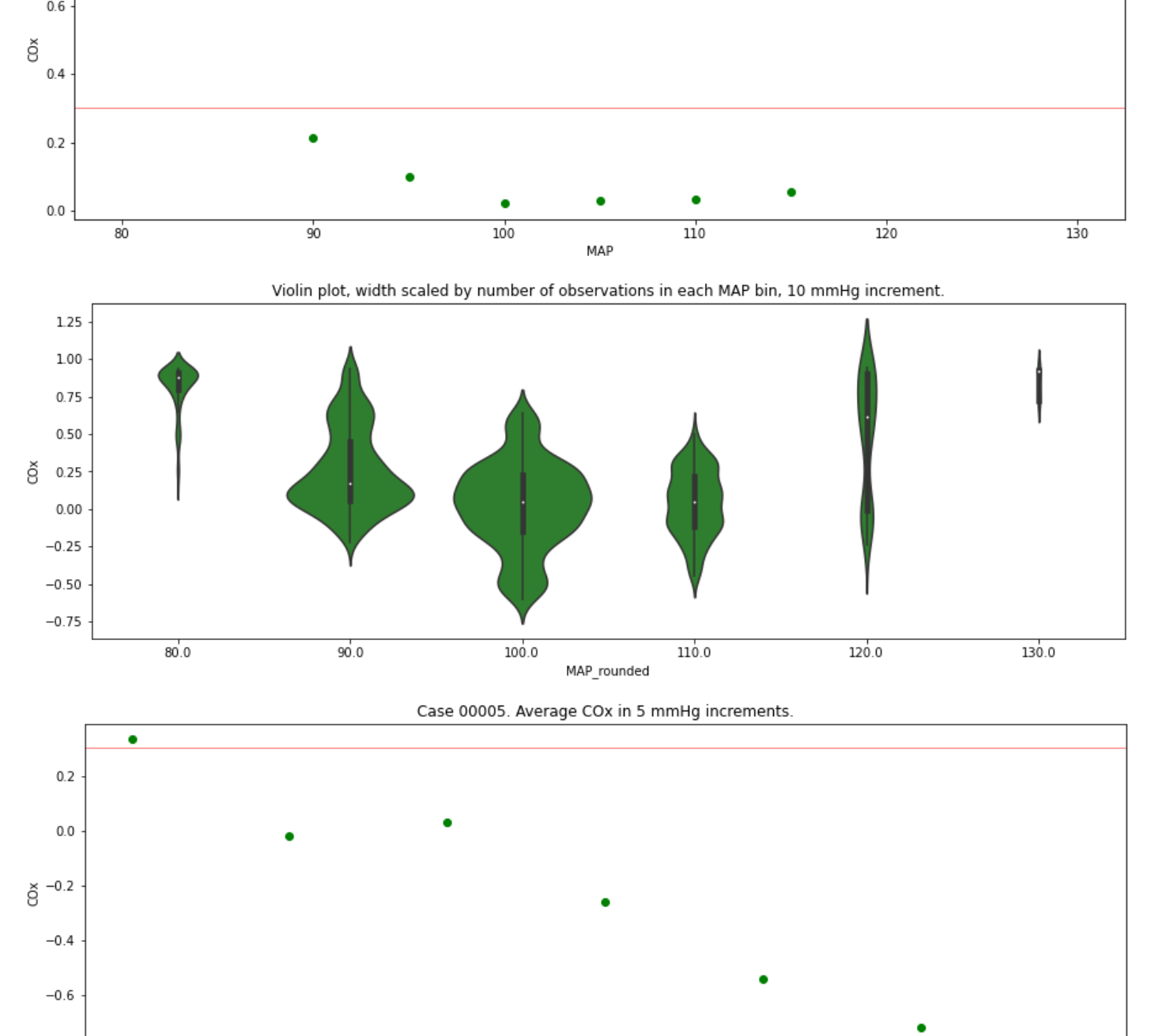
As the automatic cross-correlation method proved unreliable for aligning this type of data, time synchronization is done manually, by visual inspection.

#### Hypothesis:

Manual synchronization using visual inspection can be used to reliably align the signals.

#### Interpretation:

[View all images](#)



Effect of alignment on the MAP-rSO<sub>2</sub> pattern.

[View all images](#)

#### Conclusion:

Manual alignment was used, based on visual inspection of common patterns between the signals. Part of the cases show no clearly identifiable point of alignment.

### 6. COx index over time

For BOPRA objective "To assess the association between time on cerebral blood flow autoregulation range (defined as COx < 0.3) and mortality, morbidity and quality of life."

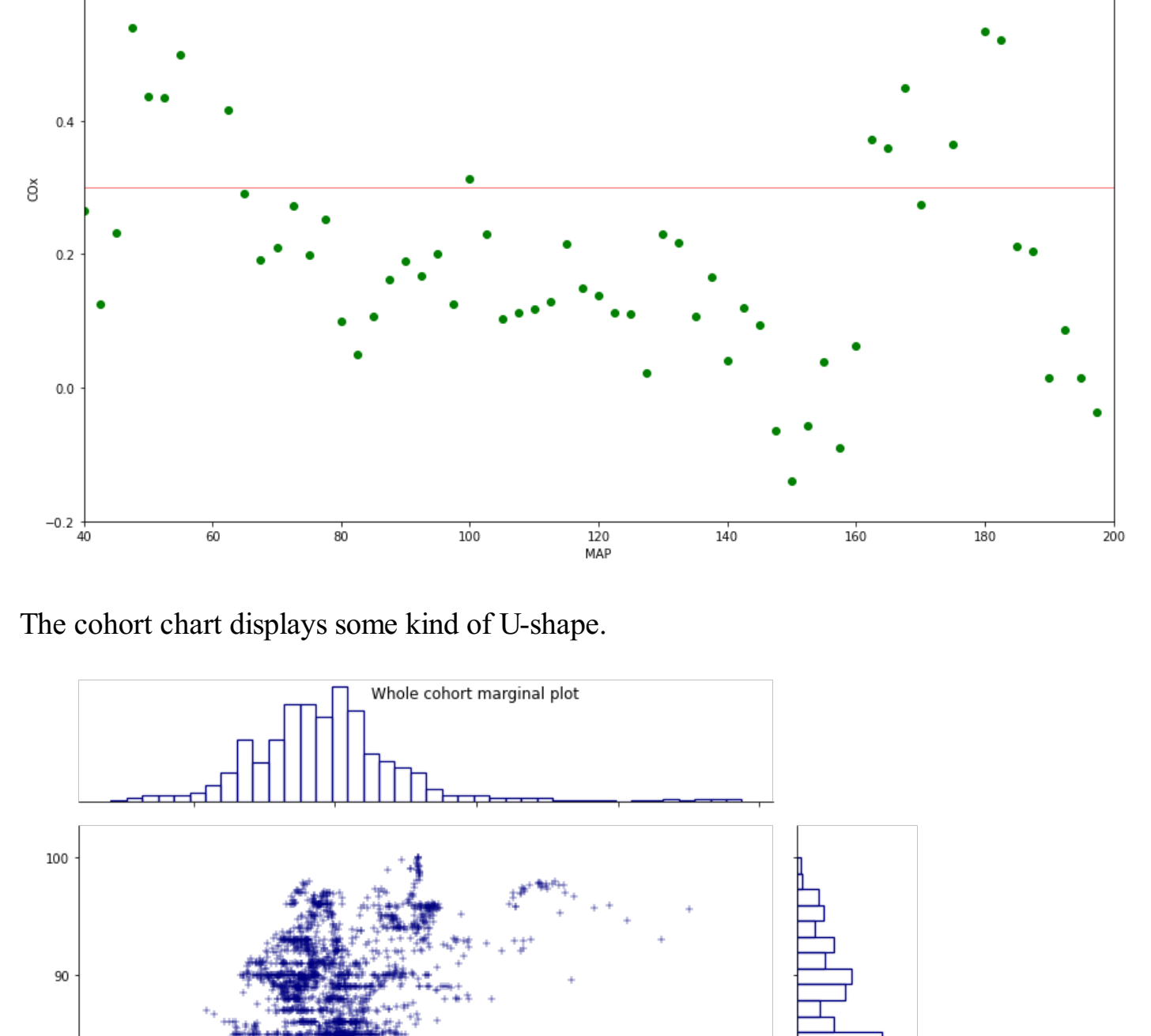
Data discontinuities will be imputed using linear interpolation.

#### Hypothesis:

The COx index can be defined for a significant duration.

#### Interpretation:

[View all images](#)



[View all images](#)

#### Conclusion:

Some figures display a u-shaped or descending curve. Violin plots were generated to see where the most of the samples fall. This helps assess the actual blood pressure variation and to detect outliers. Expert physician interpretation is required to assess the results.

### 8: Summary

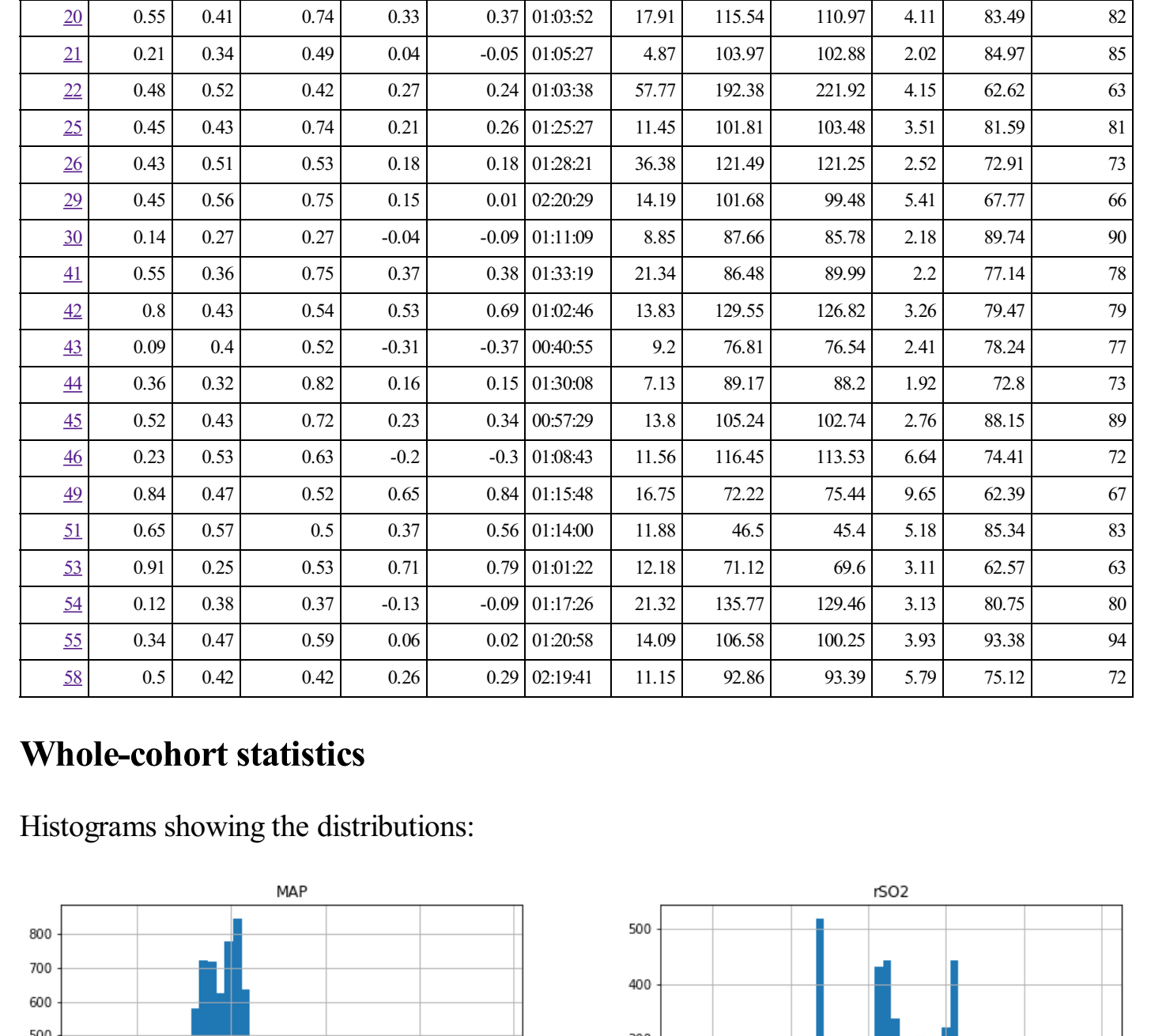
A selection of images from the previous steps were collected together. A scatterplot with marginal distributions and a fixed MAP axis was generated.

[View case summaries](#)

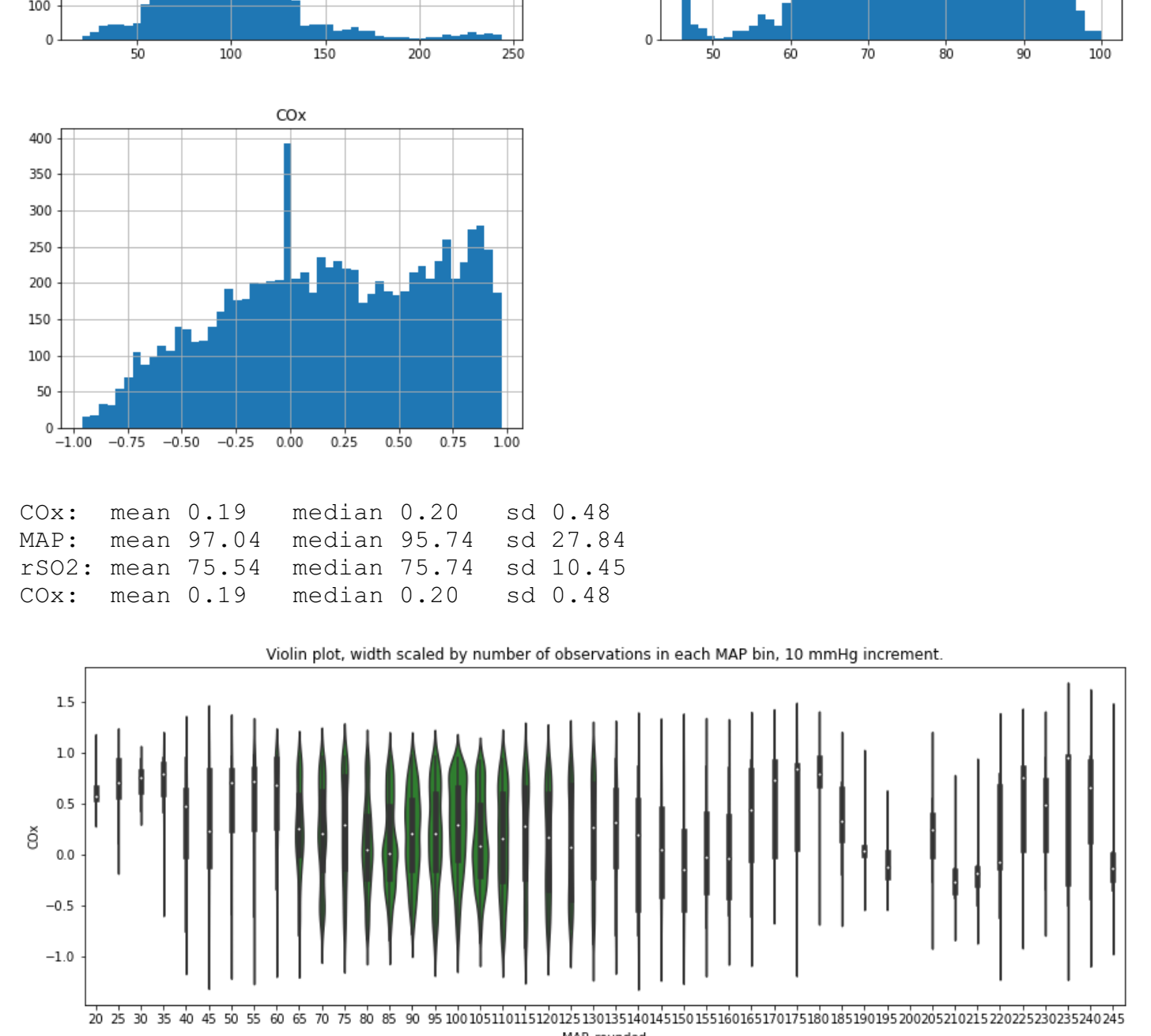
### 9: Statistics

Cohort chart and some numerical statistics were generated.

[View Jupyter notebook](#)



The cohort chart displays some kind of U-shape.



[View Jupyter notebook](#)

Case ID	COx < 0.3	COx SD	COx defined	COx mean	COx median	COx median	Duration	MAP SD	MAP mean	MAP median	rSO2 SD	rSO2 mean	rSO2 median
1	0.36	0.42	0.85	0.09	0.02	0.016913	23.92	112.08	112.65	6.25	62.5	85	
2	0.61	0.5	0.7	0.39	0.5	00:50:55	25.67	119.87	112.65	3.49	68.39	69	
3	0.45	0.53	0.58	0.18	0.21	01:38:33	19.99	88.44	81.41	7.11	75.59	72	
5	0.3	0.49	0.49	-0.06	-0.07	01:25:34	7.43	91.86	90.61	5.61	67.96	65	
6	0.54	0.34	0.62	0.37	0.37	01:20:43	20.33	103.89	90.43	4.32	94.97	96	
7	0.4	0.38	0.56	0.24	0.17	01:18:24	10.65	97.14	95.57	3.56	83.97	84	
8	0.44	0.48	0.75	0.09	0.14	01:04:03	17.55	92.63	88.06	1.27	90.13	90	
9	0.58	0.51	0.48	0.3	0.47	00:45:49	24.79	96.47	88.12	2.31	63.07	63	
10	0.25	0.35	0.73	0.06	0.13	01:38:30	4.11	68.43	68.12	5.35	66.52	64	
13	0.21	0.4	0.62	-0	-0.02	01:23:50	6.26	105.75	105.13	1.03	81.52	81	
15	0.14	0.42	0.32	-0.09	-0.09	00:42:18	7.02	75.96	76.93	7.28	62.1	58	
16	0.29	0.44	0.35	0.11	-0.02	01:22:54	16.52	84.81	80.86	2.44	83.89	83	
17	0.41	0.49	0.48	-0.23	0.24	00:52:55	5.45	69.98	68.4	5.54	51.52	49	
18	0	0.28	0.06	-0.44	-0.48	00:56:50	16.79	83.39	83.39	1.43	84.33	84	
19	0.5	0.41	0.62	0.32	0.3	01:24:41	13.32	97.6	99.5	11.77	66.97	70	
20	0.53	0.41	0.74	0.33	0.37	01:03:52	17.91	115.54	110.97	4.11	83.49	85	
21	0.21	0.34	0.49	0.04	-0.05	01:05:27	4.87	103.97	102.88	2.02	84.97	82	
22	0.48	0.52	0.42	0.27	0.4	01:03:38	57.77	192.38	121.92	4.15	62.62	63	
25	0.45	0.43	0.74	0.21	0.26	01:13:19	11.45	101.81	103.48	1.1	81.59	81	
26	0.43	0.51	0.53	0.18	0.15	01:14:00	11.88	46.5	45.4	5.18	85.34	83	
29	0.45	0.56	0.75	0.15	0.01	02:20:29	14.19	101.68	99.48	5.41	67.77	66	
30	0.14	0.27	0.27	-0.04	-0.09	01:11:09	8.85	87.66	85.78	2.18	89.74	90	
41	0.53	0.36	0.75	0.37	0.38	01:13:19	21.34	86.48	89.99	2.2	77.14	78	
42	0.38	0.43	0.54	0.53	0.69	01:02:46	13.83	129.55	126.82	2.36	79.47	79	
43	0.09	0.4	0.52	-0.31	-0.37	00:40:55	9.2	76.81	76.54	2.41	74.28	73	
44	0.36	0.32	0.82	0.16	0.15	01:30:08	7.13	89.17	88.2	1.92	72.8	75	
45	0.52	0.43	0.72	0.23	0.34	00:57:29	13.8	105.24	102.74	2.76	88.15	89	
46	0.23	0.53	0.63	-0.2	-0.3	01:08:43	11.56	116.45	113.53	6.64	74.41	72	
49	0.84	0.47	0.52	0.65	0.84	01:15:48	16.75	72.22	75.44	9.65	62.39	67	
51	0.65	0.57	0.5	0.37	0.56	01:14:00	11.88	46.5	45.4	5.18	85.34	83	
53	0.91	0.25	0.53	0.71									