



TALLINNA TEHNIKAÜLIKOOL
INSENERITEADUSKOND
Ehituse ja arhitektuuri instituut

**CO₂ ANDURITE MÕÕTEVIGADE JA
ANDMETÖÖTLUSALGORITMIDE MÕJUANALÜÜS
ÕHUKVALITEEDI JA KASUTUSAEGADE
HINDAMISEL**

**CO₂ SENSORS: IMPACT ANALYSIS OF MEASUREMENT
ERRORS AND DATA PROCESSING ALGORITHMS ON AIR
QUALITY AND OCCUPANCY ESTIMATION**

MAGISTRITÖÖ

Üliõpilane: Hans Kristjan Aljas

Üliõpilaskood: 192768EAKI
Martin Thalfeldt,
kaasprofessor tenuuris
Tuule Mall Parts,
ekspert

Tallinn 2024

AUTORIDEKLARATSIOON

Olen koostanud lõputöö iseseisvalt.

Lõputöö alusel ei ole varem kutse- või teaduskraadi või inseneridiplomit taotletud.

Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

22. mai 2024.a

Autor: *allkirjastatud digitaalselt*

/ allkiri /

Töö vastab bakalaureusetöö/magistritööle esitatud nõuetele

..... 20.....

Juhendaja:

/ allkiri /

Kaitsmisele lubatud

.....20... .

Kaitsmiskomisjoni esimees

/ nimi ja allkiri /

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Hans Kristjan Aljas,

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „CO₂ ANDURITE MÕÕTEVIGADE JA ANDMETÖÖTLUSALGORITMIDE MÕJUANALÜÜS ÕHUKVALITEEDI JA KASUTUSAEGADE HINDAMISEL“,

mille juhendaja on Martin Thalfeldt,

1.1 reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2 üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.

3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

22. mai 2024

¹ Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingu tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.

CO₂ ANDURITE MÕÕTEVIGADE JA ANDMETÖÖTLUSALGORITMIDE MÕJUANALÜÜS ÕHUKVALITEEDI JA KASUTUSAEGADE HINDAMISEL

Üliõpilane: Hans Kristjan Aljas, 192768 EAKI

Õppekava, peaariala: EAKI02/19 - Hoonete sisekliima ja veetehnika

Juhendaja(d):

kaasprofessor tenuuris / programmijuht, Martin Thalfeldt, 6202505
ekspert, Tuule Mall Parts

Lõputöö teema:

(eesti keeles) CO₂ ANDURITE MÕÕTEVIGADE JA ANDMETÖÖTLUSALGORITMIDE
MÕJUANALÜÜS ÕHUKVALITEEDI JA KASUTUSAEGADE HINDAMISEL

(inglise keeles) CO₂ SENSORS: IMPACT ANALYSIS OF MEASUREMENT ERRORS AND
DATA PROCESSING ALGORITHMS ON AIR QUALITY AND OCCUPANCY ESTIMATION

Lõputöö põhieesmärgid:

1. koguda ja analüüsida siseruumide CO₂-taseme mõõteandmeid, et hinnata ruumide õhukvaliteeti ja kasutusaegasid täienevate üle-euroopaliste sisekliimanõuete valguses;
2. suurendada teadlikkust mõõtetulemusi mõjutavatest teguritest ja lahendustest, et saavutada võimalikult tõetruu andmestik;
3. arendada välja ja kalibreerida laialdaselt rakendatavad andmetöötlusmeetodid levinud mõõtevigade korrigeerimiseks ja ruumide kasutusaegade hindamiseks;
4. kvantifitseerida mõõtevigade, andmete järeltöötlusmeetodite ja monitoorimist pärssivate olmeprobleemide mõju tulemustele.

Lõputöö etapid ja ajakava:

Nr	Ülesande kirjeldus	Tähtaeg (2024)
1	Tööks vajalike mõõteandmete allikate leidmine ja andmete esmane ülevaatus.	1. veebr
2	Töö eesmärkide täpne kirjeldus. Esmaversioonid: kirjanduse ülevaatus, teoreetilised alused, töö metoodika struktuur ja etappide lühikirjeldused.	26.veebr
3	Tööks vajalike mõõteandmete kogumine, kategoriseerimine ja põhjalik ülevaatus analüüsiks.	4.märts
4	Metoodika punktide 4.1 ja 4.2 (esmane kvaliteedikontroll ja probleemide liigitamine) tööprotsess ja tulemuste esmane kirjeldus.	11.märts
5	Metoodika punkti 4.3 (baastaseme korrektsioonialgoritm) tööprotsess ja tulemuste kirjeldus.	18.märts
6	Baastaseme korrektsioonialgoritmi valideerimine ja tulemuste kirjeldus.	25.märts
7	Metoodika punkti 4.4 (kasutusaegade tuvastamine) tööprotsess ja tulemuste kirjeldus. Täiendavate mõõteandmete allalaadimine ja ülearvutamine.	1.aprill
8	Kirjanduse ülevaatus, teoreetiliste aluste, metoodika kirjelduse ja tulemuste lõplik vormistamine, vastavusse viimine viimaste tulemuste ja läbiviidud tegeliku tööprotsessiga.	15.aprill
9	Järeldused. Kokkuvõtte. Kasutatud kirjandus. Graafikute kujundus, vormistus, stiili- ja keelekontroll, detailide parandamine.	22.aprill
10	Lõputöö lõplik esitamine juhendajatele. Viimased parandused. Puhveraeg tekkinud ettenägematute probleemide lahendamiseks.	7.mai

Töö keel: eesti keel**Lõputöö esitamise tähtaeg:** 22. mai 2024.a

Üliõpilane: Hans Kristjan Aljas *allkirjastatud digitaalselt* 22. mai 2024.a
/allkiri/

Juhendaja: Martin Thalfeldt20.....a
/allkiri/

Juhendaja: Tuule Mall Parts20.....a
/allkiri/

Programmijuht: Martin Thalfeldt20.....a
/allkiri/

SISUKORD

EESSÕNA	8
Lühendite ja tähiste loetelu	9
1. SISSEJUHATUS.....	10
2. KIRJANDUSE ÜLEVAADE	11
3. TEOREETILISED ALUSED	17
3.1 Euroopa hoonete energiatõhususe direktiiv	17
3.2 CO ₂ sisekliima indikaatorgaasina.....	17
3.3 CO ₂ põhine kasutusaegade hindamine ja nõudluspõhine ventilatsioon	18
3.4 CO ₂ andurite NDIR tehnoloogia.....	18
3.5 EN 16798-1:2019 ja TAIL metoodika rakendamine õhukvaliteedi hindamiseks.	19
4. METOODIKA.....	21
4.1 Mõõteandmete kogumisprotsessid	22
4.2 Mõõteandmete puhastamine ning andmekadu	22
4.3 Baastaseme korrektsioon.....	23
4.3.1 Algoritmi tööpõhimõte	23
4.3.2 Parameetrite optimeerimine ja algoritmi mõjuanalüüs.....	24
4.4 Kasutusaegade tuvastamine.....	29
4.4.1 Algoritmi tööpõhimõte	29
4.4.2 Tundlikkuse analüüs	31
4.4.3 Mõju sisekliimaklassidesse jaotumisele	32
4.5 Töö limitatsioonid	33
5. Mõõteandmete puhastamine ning andmekadu	34
5.1 Mõõteandmete puhastusetapid	34
5.1.1 Duplikaadid ja valemärgistused.....	34
5.1.2 Mõõteandmete pidevus	34
5.1.3 Salvestamise intervall ja <i>resampling</i>	35
5.1.4 Nullväärtuste eemaldamine	36
5.1.5 Stagneerunud väärtuste eemaldamine	36
5.1.6 Skaalast väljasolevate väärtuste eemaldamine	37
5.1.7 Lühikeste andmeaukude täitmine	37
5.1.8 Järelejäänud anomaaliad	38
6. TULEMUSED JA ANALÜÜS	39
6.1 Mõõteandmete puhastamine ning andmekadu	39
6.1.1 Andmekao kvantifitseerimine	39
6.1.2 Analüüs	39
6.2 Baastaseme korrektsioon.....	40

6.2.1 Hajuvusdiagrammidega tutvumine	40
6.2.2 Hajuvusdiagrammide piirkondade analüüs	43
6.2.3 Optimaalsete parameetrite valik	49
6.2.4 Optimeeritud parameetritega korrektsioonialgoritmi mõju tulemustele	53
6.3 Kasutusaegade tuvastamine	56
6.3.1 Algoritmi töö ülevaade ja piirangud	56
6.3.2 Tuvastusstsenaariumite võrdlus	57
7. Järeldused	61
KOKKUVÕTE	64
SUMMARY	65
KASUTATUD KIRJANDUSE LOETELU	66
LISA 1 - Baastaseme korrektsiooni valideerimine piirkontsentratsioonide meetodil	69

EESSÕNA

Käesolev töö on tugineb FinEst Targa Linna Tippkeskuse pilootprojektil DigiAudit, mis loodi eesmärgil hõlbustada hoonete energiatõhususe ja sisekliima reaalaajalist jälgimist digitaalse platvormi vahendusel. DigiAuditi platvormi tehnilise lahenduse, sealhulgas andmete kogumise teostas Thinnect OÜ.

Magistritöö sündis Martin Thalfeldti initsiatiivist uurida edasi DigiAuditi käigus kogutud CO₂ mõõteandmeid Euroopa Hoonete Energiatõhususe Direktiivi (EPBD) uuenduste valguses. Tänan Martin Thalfeldti suunamise ja kontseptsioonide eest ning Tuule Mall Partsi vahetu tagasiside, heade nõuannete ja praktilise abi eest. Samuti tänan Jarek Kurnitskit, kes jagas ajakohast informatsiooni EPBD uuenduste kohta. Töö valmimisel oli vahetuks abiliseks ka ChatGPT, kes/mis hõlbustas uute ideede ja matemaatiliste kontseptsioonide avastamist, programmeerimise tööprotsessi ning andmete visualiseerimist.

Magistritöö märksõnad: CO₂ andurid, sisekliimaklass, kasutusaegade tuvastamine, EN 16798-1:2019, magistritöö

Lühendite ja tähiste loetelu

BLC – baastaseme korrigeerimine (*baseline correction*)

BMS – digitaalne hoonehaldussüsteem (*Building Management System*)

DCV – nõudluspõhine ventilatsioon (*demand controlled ventilation*)

EPBD – Euroopa hoonete energiatõhususe direktiiv (*Energy Performance of Buildings Directive*)

ICC – arvutuslik sisekliimaklass (*Indoor Climate Class*)

NDIR – levinud anduritehnoloogia gaaside kontsentratsioonide tuvastamiseks proportsionaalselt infrapunakiirguse neeldumisintensiivsusele (*non-dispersive infrared*)

REHVA – Euroopa kütte- ja ventilatsiooniühingute esindajate föderatsioon (*Representatives of European Heating and Ventilation Associations*)

Resampling – aegrea sissekannete ajaintervalli ühtlustamine valitud sagedusele.

Mõõteväärtused arvutatakse vastavalt ümber valitud meetodika alusel.

1. SISSEJUHATUS

Siseruumides veedetakse arenenud riikides ligi 90% ajast (Klepeis et al., 2001). Hoonete sisekliima kvaliteedil on oluline mõju ruumisviibijate tervisele ja kognitiivsele sooritusvõimekusele, seega on oluline tagada head sisekliimatingimused, et ei halveneks õppe- ja töövõimekus (Stafford, 2015) ning et viiruste õhukaudse leviku risk oleks minimeeritud (Peng et al., 2022). Sisekliima kvaliteeti ja sellest tulenevalt ka tehnosüsteemide eesmärgipärast toimimist saab süstemaatiliselt hinnata, monitoorides ja analüüsides kasutuses olevate ruumide temperatuuri ja süsihappegaasi (CO₂) jooksivaid mõõteandmeid. CO₂ taseme põhjal on võimalik hinnata ka ruumi kasutusprofiili ja sellest tulenevalt optimeerida ventilatsiooni ja kütte- või jahutusseadmete tööd või lausa ruumide kasutusotstarve ümber mõtestada. Lisaks on ruumiandurite abil jälgitav õhukvaliteet heaks esmaseks indikaatoriks, et tuvastada puudulikult töötavaid ventilatsioonisüsteeme või probleemseid üksikuid ruume, vältides keerukamaid lahendusi nagu andurite lisamist õhukanalitesse. Hea sisekliima tagamine ja jälgimine on osa ka uuenenud 2024. a Euroopa hoonete energiatõhususe direktiivis (EPBD) liikmesriikidele esitatavatest suunistest.

Eelnev tõstatab kõik sisekliima andmepõhist jälgimist ja hindamist puudutavad teemad päevakorraliseks, kuna direktiivi jõustumisel tuleks hakata koguma ja analüüsima sisekliimat iseloomustavaid mõõteandmeid suuremal skaalal, samas on varasemalt täheldatud laialt kasutuses olevate CO₂ andurite ebastabiilset toimimist, sh märkimisväärseid mõõtevigasid (Mylonas et al., 2019). Kurnitski (REHVA) mainib EPBD uuendust selgitavas artiklis, et kuigi CO₂ ruumiandurid on osana nõudluspõhistest ventilatsioonisüsteemidest juba turul olnud pikka aega, on nende süsteemide paigalduskvaliteet olnud sageli puudulik ning süsteemid tervikuna ei ole tõestanud oma töökindlust (Kurnitski, 2023). Sellest tulenevalt on ruumipõhiste CO₂ andurite kasutamine siiani suhteliselt vähelevinud ning mõõtmisi pärssivate tegurite ja mõõtevigade mõju vajab kvantifitseerimist ning lihtsaid lahendusi vigade eemaldamiseks.

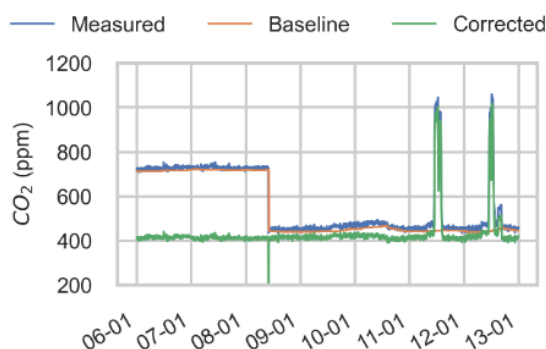
Antud töö eesmärk on esitada statistiline taustinfosüsteem uuenevatele sisekliima hindamismetoodikatele ja suunistele ning kasvavale CO₂-andurite kasutustrendile, et suurendada teadlikkust anduritega seotud põhiprobleemidest, nende mõjust mõõtetulemustele ning võimalikest kompensatsiooni- ja järeltöötlusmeetmetest.

Töö käigus hinnatakse mõõteandmete kättesaadavust ja kvaliteeti, seejärel pakutakse välja järeltöötusalgoritmid ja optimeeritakse nende tööd. Võrreldakse töödeldud ja töötlemata andmete mõju tulemustele lähtuvalt lähtuvalt Euroopa standardi EN 16798-1:2019 hindamismetoodikast.

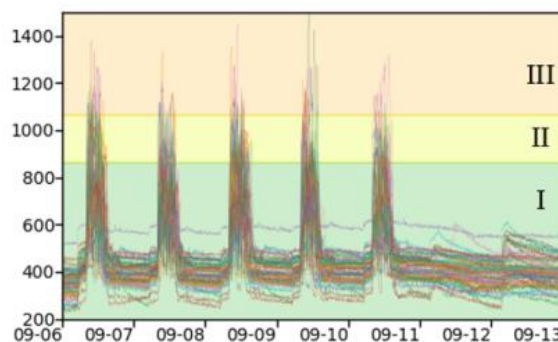
2. KIRJANDUSE ÜLEVAADE

CO₂ andurite täpsus

Vajadus pöörata tähelepanu CO₂ andurite näitude korrektsusele avaldus, kui antud töö autor analüüsis 56 klassiruumi aastaseid CO₂ mõõteandmeid ja ilmnedid süstemaatilised vead ning anomaalsed hüpped (Joonis 2.1) mõõteandmete baasnivoos. Mõõtmistes kasutatud andurite deklareeritud täpsus oli 25°C juures ± 30 ppm. Lisaks oli anduritele sisseehitatud automaatne baasnivoo korrigeerimisalgoritm, aga mõõteandmetest avaldusid kohati baasnivood isegi 250 ppm juures, mis on atmosfääri süsihappekontsentratsiooni (≈ 420 ppm) arvestades võimatu ja tähendab minimaalselt 150 ppm suurust näidu alahindamist (Aljas et al., 2023). Ligikaudu sama magnituudiga ilmnedid ka baasnivoo ülehindamise vead (Joonis 2.2). Jooniselt nähtub, et isegi nädalavahetusel, kui õpperuumid kasutuses polnud, olid CO₂ mõõdetud tasakaalukontsentratsioonid andurite lõikes väga erinevad. See tõdemus annab kindla indikatsiooni, et mõõteväärtused on nihkunud ja ei kajasta tegelikku olukorda eeldatava täpsusega.



Joonis 2.1 Anomaalne hüpe mõõdetud CO₂ tasemes (Vihman et al., 2023)



Joonis 2.2 56 CO₂ anduri 7 päeva mõõteandmed demonstreerimaks baastaseme varieeruvust (Aljas et al., 2023)

Eelnimetatud uurimistöös kasutati NDIR (*non-dispersive infrared*) tüüpi CO₂ andureid, mis on hoonete tehnosüsteemides kasutatavate CO₂ anduritena enim levinud anduritüüp, taskukohase hinnaga ja deklaratsiooni järgi ka piisava täpsusega. Samas leidub palju teaduskirjandust, kus vastava tehnoloogiaga andurite puhul on märgatud tugevaid hälbeid väljaspool deklareeritud täpsusklassi ja sarnaseid probleeme eelneva uuringuga.

Fisk et al. uurisid 44 NDIR CO₂ andurit üheksas avalikus hoones. Täheldati kuni mitmesaja ppm suuruseid mõõtevigasid ja suurt variatsiooni, millest järeldati vajadust kvaliteetsemate andurite ja parema kalibreerimise järele (W. Fisk et al., 2008).

Fisk et al. täiendasid oma tööd, uurides 208 NDIR CO₂ andurit 34 ärihoones. Täheldati suurt varieeruvust andurite täpsuses lähtuvalt tootjafirmade lõikes, kuid isegi

kõrgemakvaliteedilised andurid ei taganud kohalikule standardile vastavat täpsusklassi (W. J. Fisk et al., 2010). Mõõtevigade peamiseks põhjuseks toodi otsesed probleemid anduriga ja vähemtõenäolisemaks peeti tõrkeid andmeedastusahelas või anduri vananemist. Raporti hinnangul on nõudluspõhise ventilatsiooni (DCV) efektiivsuse tagamiseks vaja arengut andurite tehnoloogias.

Shrestha ja Maxwell teostasid 4-osalise uuringu hindamaks laialtlevinud NDIR CO₂ andurite tööd hoonete tehnosüsteemide kontekstis. Esimeses uuringus hinnati andurite täpsust, lineaarsust, korratavust, hüstereesi ja tundlikkust õhuniiskuse, -temperatuuri ja -rõhu suhtes. Sama mudeliga andurite mõõtetulemustes leiti märkimisväärseid erinevusi (Shrestha & Maxwell, 2009). Uuringusarja teises osas analüüsiti 45 andurit 15 erinevast mudelitüübist ja täheldati olulisi erinevusi andurite deklareeritud ja tegelike täpsuste vahel kõigi tootjate puhul, kusjuures ükski andur ei vastanud deklareeritud täpsusklassile (Shrestha & Maxwell, 2010a). Kolmandas osas uuriti andurite tundlikkust keskkonnatingimustele. Niiskuse suhtes tuvastati tundlikkus vaid kolmel anduril. Leiti, et õhutemperatuur ja -rõhk mõjutasid mõõtmistulemusi kõigi andurite puhul, kusjuures erinevused avaldusid jällegi ka samade mudelite lõikes. Samuti märgiti ära kalibreeringu tugev mõju tulemustele (Shrestha & Maxwell, 2010b). Uuringu neljas osa keskendus anduri vananemise mõjule. Leiti, et aastase katseperioodi jooksul avaldus vananemise mõju mõne anduritüübi puhul märgatavalt, mõne puhul üldse mitte (Shrestha & Maxwell, 2010c).

2018. aastal teostasid analoogseid uuringuid Mylonas et al., kasutades kuute erinevat NDIR CO₂ andurit (Mylonas et al., 2019). Vaatamata ligi 8-aastasele arenguperioodile võrreldes elnevate uuringutega, ilmnisid endiselt sarnased probleemid, kuid sõltuvalt mõjuteguritest käitusid andurid erinevalt. Kolme anduri mõõtevigaga kasvas temperatuuri suurenedes, kolmel CO₂ kontsentratsiooni suurenedes ja ühel suhtelise niiskuse suurenedes. Siiski ei saavutanud ükski andur deklareeritud mõõtetäpsust – vead standardsetel töötingimustel ulatusid kuni 400 ppm-ni, olles valdavalt suuremad kui 100 ppm.

Ka Berquist ja Banister hoiatasid, et andurite pimesi kasutamine ilma valideerimata võib nõudluspõhise ventilatsioonisüsteemi töö rikkuda: uuritud kolmest andurist üks näitas algselt väärtuseid nihkega 500 ppm, ühel anduril olid näidud 19% ajast lubatud täpsuspiirkonnast väljas ning kolmas andur ei edastanud andmeid üldse (Berquist & Banister, 2020).

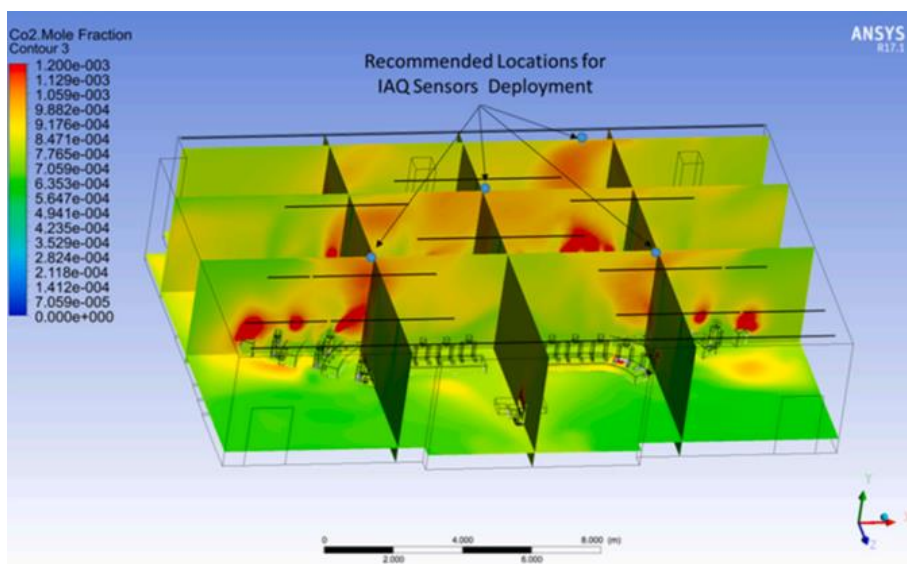
Annaliese Keimel täheldas samuti ligi 70...100 ppm suurust nihet mõõteandmetes, kuid tões, et korrektsioonimeetmeid rakendades ühtis dūnaamika referentsandmetega hästi (Keimel, 2019).

Baastaseme korrektsiooni algoritm

Vihman et. al on pakkunud välja lihtsa reaalajas toimiva algoritmi, et parandada ajas nihkuvat mõõdetud CO₂ baastaset adekvaatselt ventileeritud perioodilise kasutusega ruumides (Vihman et al., 2023). Meetod põhineb üldistusel, et ruumikasutajate puudumisest ja õhuvahetusest tingituna saavutatakse liikuva ajaakna raamides CO₂ taseme baasnivoo ehk välisõhu CO₂ kontsentratsioon, mille järgi teostatakse korrektsioonid vastavalt tegelikule hinnangulisele välisõhu CO₂ tasemele. Eeldatakse ka, et baasnivoost madalamad erandid (*outliers*) mahuvad alumise 1-protsentiili hulka. Töös väljapakutud 9-tunnine ajaaken põhineb vajadusel vähendada algoritmi viibeaga reaalajas kasutusolukorras, kuid autor nendib, et pikemad ajaaknad annavad stabiilsema väljundi ning ajaakna suuruste valiku mõju tuleks täiendavalt uurida.

CO₂ andurite paigutus

Mou et al. hindasid CFD simulatsioone põhjal CO₂ andurite paigutuse mõju mõõtetulemustele 210 m² suuruses seguneva õhuvahetusega auditoriumis (Mou et al., 2022). Inimeste vahetus läheduses oli CO₂ tase ligi 1200 ppm ning ülejäänud ruumis varieerus 600...1000 ppm vahel. Autorid pakkusid välja optimaalseteks asukohtadeks spetsiifilised laealused punktid 10...30 cm vahega laest. Samas paistab jooniselt, et väljapakutud asukohtade ümbruses on kontsentratsiooni gradient endiselt suur (Joonis 2.3). Viibimistsoonis seinte lähistel, kuhu andurid sageli tegelikult paigutatakse, oli kontsentratsioon jooniselt hinnatuna 200...400 ppm võrra madalam.



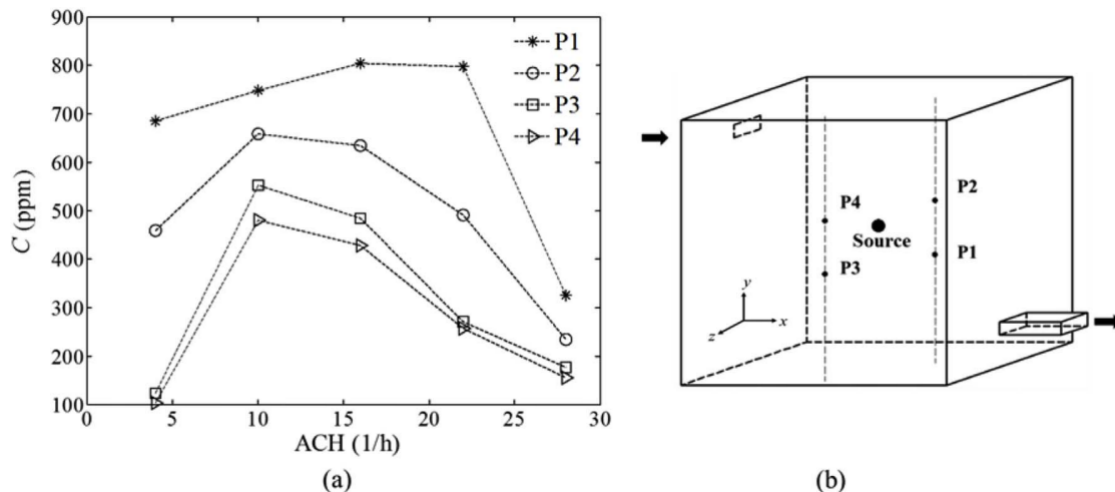
Joonis 2.3 CO₂ kontsentratsioonide jaotus ja soovituslikud paigutusasukohad auditoriumis (Mou et al., 2022)

Mahyuddin ja Awbi katsetasid CO₂ jaotumist väikeses, 2 istuva inimese ekvivalendiga testkambris (Mahyuddin & Awbi, 2010). Märgiti ära, et kuigi oma suurema tiheduse tõttu peaks süsihappegaas alla vajuma, oli domineerivaks teguriks gaasi levikul sooja

õhu konvektsioon, mis kandis gaasi just ruumi kõrgematesse kihtidesse. Kõrgeima CO₂ kontsentratsiooni piirkonnad ühtisid kõrgeima temperatuuri piirkondadega ruumis. Siiski jäid mõõdetud erinevused ventileeritud olukordades ± 50 ppm suurusesse vahemikku.

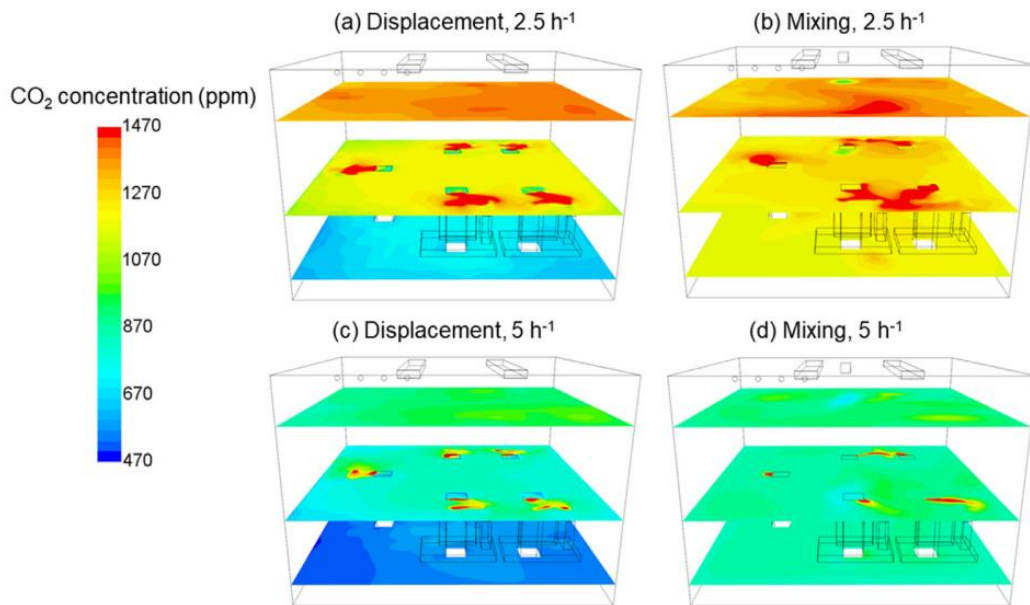
CO₂ kihistumist 3 x 3 x 3 m testkambris uurisid ka Deng et al. (Deng et al., 2018). Nad märkisid, et sissepuhkekanalist väljuva joa keeriselisus (*vorticity*) on olulise mõjuga õhu efektiivsel segunemisel. Isegi suurtel õhuvahetuskordsustel esines suure sissepuhkeava puhul märgatav kihistumine madalama joa kiiruse ja väiksema keeriselisuse tõttu.

Esinesid äärmiselt suured erinevused (kuni 600 ppm) mõõdetud CO₂ tasemetes lähestikustes asukohtades saasteallika ümbruses. Samas kasutati NDIR andureid, mille kalibreerimist ei mainitud. Sellest tulenevalt on töö tulemuste tõlgendamine praktilises kontekstis keeruline.



Joonis 2.4 CO₂ kontsentratsioonide varieerumine erinevate mõõtepunktide (P1 kuni P4) ja õhuvahetuskordsuste (ACH) puhul (Deng et al., 2018)

Pei et al. tegid samuti katseid kliimakambris (Pei et al., 2019). Kamber oli mõõtmetega 4,3 x 4,3 x 4,3 m ning sisaldas 5 soojus- ja CO₂ eraldusega mannekeeni. Tõrjuva õhuvahetuse puhul, õhuvahetuskordsusel 2,5 h⁻¹ märgati 2,4 m mõõtmiskõrgusel 200 ppm võrra kõrgemat näitu kui 1,2 m kõrgusel. Õhuvahetuse suurendades 5 h⁻¹ peale vähenes vastav erinevus 100 ppm-ni. Seguneva õhuvahetusega olid erinevused vastavalt alla 150 ppm ja alla 50 ppm. Samas säilisid konvektiivsete õhuvoolude kantud kõrgeimad kontsentratsioonid piirkonnad.



Joonis 2.5 CO₂ kontsentratsioonide jaotumine ruumis erinevatel ventilatsioonirežiimidel (Pei et al., 2019)

Soojusliku konvektsiooni tähtsuse märkisid ära ka Bulińska et al., uurides kalibreeritud CFD mudeliga magamistuba, kus oli loomulik ventilatsioon, üks magav inimene, avatud aken ja radiaator. Autorid leidsid, et vältida tuleks soojus- ja saasteallikate (sh inimeste) vahetus läheduses ja kohal asuvat ruumpiirkonda, kuna seal on konvektiivsete jugade mõju suur ja õhk pole piisavalt segunenud (Bulińska et al., 2014).

Kõik eelmainitud uuringud CO₂ jaotumise kohta ruumides kirjeldavad nähtuse füüsikalist poolt sarnaselt, samas on andurite asukohavalikul soovitud mõningaselt varieeruvad.

REHVA on samuti jaganud soovitusi CO₂ andurite paigaldamiseks sisekliima hindamisel (tõlgituna): „Ideaalne anduri asukoht on inimeste peamistele statsionaarsetele viibimiskohtadele võimalikult lähedal. Kontorihoones tähendaks see näiteks andurite paigutamist töölaudadele, 1...1,2 m kõrgusele põrandast. Kui see ei ole võimalik, siis peaks paigutama anduri lähedalasuvale seinale (näiteks termostaadi kõrvale). Paremusest kolmas asukoht oleks lae all, kuid (ripp)lakke või ventilatsioonikanalitesse paigutamist tuleks vältida, kuna see ei anna õiget infot viibimistsooni tajutavale õhukvaliteedile, välja arvatud siis, kui eesmärgiks on mõõta hoone tehnosüsteemide tööd.“ (Boerstra et al., 2019).

Kasutusaegade tuvastuse algoritm

Käesoleva töö autor rõhutas oma eelnevas uurimustöös ruumi kasutusaegade hindamise täpsuse tähtsust EN 16798-1:2019 rakendamisel, demonstreerides ruumide mõõtetulemuste jaotumist EN sisekliimaklassidesse kolmes olukorras: ilma kasutusaegade määramiseta, ruumitüüpide kontsantsete kasutusaegadega ning CO₂ kontsentratsiooni põhise dünaamilise meetodiga (Aljas et al., 2023). CO₂ kontsentratsiooni põhine meetod tugineb teadmisel, et süsihappegaasi peamiseks allikaks siseruumides on seal viibivad inimesed, seega tähendab tõusev või teatud karakteristikuga kõrgenenud CO₂ tase ruumisviibijate olemasolu.

CO₂-põhiseid kasutusaegade tuvastusalgoritme on välja pakutud mitmeid, kuid nende rakendamine nõudab sageli tehisnärvivõrkude (ANN) või teiste masinõppemeetodite kasutamist koos märgistatud treeningandmetega. Autor arendas Pedersen et al. meetodist (Pedersen et al., 2017) inspireeritult välja uue algoritmi, kuna olemasolev meetod oli keerukas ning nõudis kõrget mõõteandmete resolutsiooni ning sisaldas mitmeid empiirilisi konstante.

Antud töö uudsus seisneb

- kirjanduses esitatud anduriprobleemide mõju hindamises laia ja mitmekesise andmehulga põhjal pärast mitmeaastast arenguperioodi andurite tehnoloogias;
- suurel skaalal teostatud CO₂ andmete kogumisel avaldunud täiendavate probleemide teadvustamises, rühmitamises ja mõju hindamises;
- lihtsate ja skaleeritavate korrektsiooni- ja kasutusaegade tuvastusalgoritmide väljapakumises, nende töö ning parameetrite optimeerimises ja piiritlemises;
- erinevate stsenaariumite mõju kvantifitseerimises vastavalt Euroopa Liidus tunnustatud ning üha enam kasutust leidvatele sisekliima hindamismetoodikatele ja -nõuetele (EN 16798-1:2019, LEVEL(s), EPBD).

3. TEOREETILISED ALUSED

3.1 Euroopa hoonete energiatõhususe direktiiv

Euroopa hoonete energiatõhususe direktiiv (EPBD) aastast 2018 mainib artiklis 5, et hoonete energiatõhususe eesmärkide saavutamisel ei tohi halveneda sisekliimakvaliteet. Direktiivi uuenenud versioonis artiklis 2, 57a mõtestatakse „sisekliimakvaliteet“ lahti kui hinnang parameetritele nagu ruumi õhutemperatuur, suhteline niiskus, ventilatsiooni õhuvahetuskordsus ja erinevate saasteainete olemasolu, mis mõjutavad ruumisviibijate tervist ja heaolu. Lisaks mainitakse artiklites 7 ja 8, et uute ja oluliselt renoveeritavate hoonete puhul tuleb pöörata tähelepanu optimaalset sisekliimakvaliteeti puudutavatele teemadele. EPBD uue versiooni artikkel 11 kohustab liikmesriike sätestama nõudeid adekvaatsete sisekliimakvaliteedi standardite rakendamiseks, et tagada tervislik sisekliima. Samuti oodatakse, et liikmesriigid nõuaks uute ja oluliselt renoveeritavate mittelehoonete varustamist monitoorimis- ja juhtseadmetega asjakohaste üksuste tasemel, et tagada pideva kasutusega ruumide sisekliimakvaliteet. Uue EPBD artikli 11 2a, 3 järgi võivad liikmesriigid esitada selliseid nõudeid soovi korral ka eluhoonetele.

J. Kurnitski (REHVA) tõdeb EPBD uuendusi lahtimõtestavas artiklis, et EPBD regulatsioon ei anna täpsemaid juhiseid, kuidas läheneda sisekliima monitoorimisele (Kurnitski, 2023). „Monitoorimis- ja juhtseadmed asjakohaste üksuste tasemel“ on Kurnitski artiklis tõlgendatud kui „nõudluspõhine ventilatsioon ja temperatuuri- ning CO₂ andurid“, viidates siinkohal Euroopa jätkusuutlike hoonete raamistikule LEVEL(s), kus kasutatakse indikaatoritena justnimelt õhutemperatuuri ja CO₂ mõõteandmeid. LEVEL(s) meetodika tugineb omakorda Euroopa standardil EN 16798-1:2019.

3.2 CO₂ sisekliima indikaatorgaasina

Uuringud on näidanud, et CO₂ kontsentratsioon siseruumides on heas korrelatsioonis lenduvate orgaaniliste ühendite (VOC) ja teiste saasteainetega. Sellest tulenevalt peetakse süsinikdioksiidi heaks ruumiõhu kvaliteedi indikaatoriks.

Ruumi CO₂ tase kujuneb CO₂ eraldus- ning eemaldusvõimsuste tasakaalu järgi. Kõrgenenud CO₂ kontsentratsioon viitab ebapiisavale õhuvahetusele, mõjudes negatiivselt inimeste kognitiivsele võimekusele, healule ning tervisele. Üks märkimisväärne terviserisk siseruumides on viiruste õhukaudne levik. Inimeste eritatud aerosoolsed osakesed aitavad viiruseid edasi kanda ning seega tuleks ruumist võimalikult kiirelt välja ventileerida.

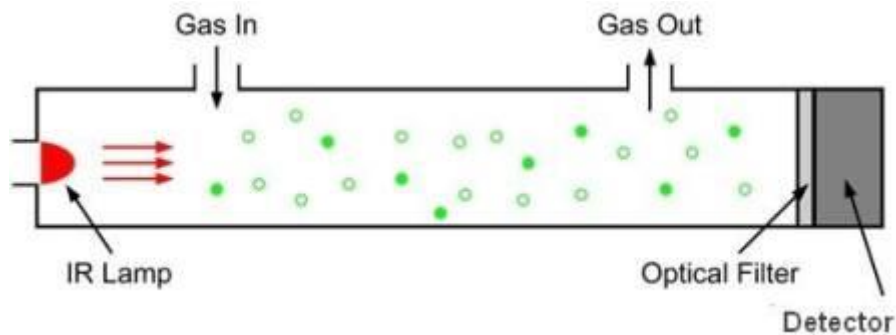
3.3 CO₂ põhine kasutusaegade hindamine ja nõudluspõhine ventilatsioon

CO₂ on inimese hingamise kõrvalsaadus ja selle kontsentratsioon võib kiiresti tõusta, kui ruumisviibijaid on palju ning õhuvahetus pole piisav. CO₂ taseme järgi saab hinnata inimeste kohalolu ning seeläbi hoone või ruumide kasutusprofiile. Kasutusprofiilide teadmine aitab optimeerida hoone tehnosüsteemide tööd energiasäästu saavutamiseks. Näiteks kasutatakse CO₂ taset sageli nõudluspõhise ventilatsiooni (DCV) juhtimiseks, et vältida ebavajalikku ventileerimist madala kasutuskooormusega ajaperioodidel.

3.4 CO₂ andurite NDIR tehnoloogia

NDIR tüüpi andurid on hoonete tehnosüsteemides enimlevinud tänu taskukohasele hinnale ja suhteliselt heale (deklareeritud) mõõtetäpsusele. NDIR andurite laiem kasutus hõlmab erinevate gaaside, sh CO₂ kontsentratsioonide tuvastamist.

NDIR andurite töö põhineb faktil, et erinevad gaasid neelavad infrapunaspiktrumis erinevaid sagedusvahemikke. Anduris paiknev valgusallikas eritab proovikambrisse vastavalt mõõdetavale gaasile iseloomuliku lainepikkusega infrapunakiirgust. Kambri teises otsas paikneb optiline sensor, mis mõõdab saabunud kiirguse intensiivsust ja selle põhjal arvutab välja gaasi kontsentratsiooni – mida rohkem kiirgust on neeldunud, seda suurem on kontsentratsioon. Anduri ees paikneb tavaliselt optiline selektiivfilter, mis aitab eemaldada spektraalset müra ja teistest gaasidest tulenevat interferentsi. Interferentsi kirjeldavad ka Dinh et al., kuid tõdevad, et veeaurust tingitud spektraalne interferents on siiani vaid osaliselt lahendatud probleem (Dinh et al., 2016). Aastal 2022 on Teng ja Chen pakkunud probleemile välja matemaatilise kompensatsiooni-mehhanismi tuginedes suhtelise niiskuse mõõteandmetele (Teng & Chen, 2023). Hoonete tehnosüsteemide NDIR CO₂ andurites kasutatakse peamiselt kahte erinevat tüüpi optilisi sensoreid – termopiile (enamlevinud) või püroelektrilisi detektoreid (kiirema reaktsiooniajaga). Mõlemate puhul on täheldatud sõltuvust temperatuurist ja teistest keskkonnateguritest, kuid samuti on arendatud välja mitmeid kompensatsioonimeetmeid (Park et al., 2010) (Yi, 2017) (Zhou et al., 2021).



Joonis 3.1 NDIR anduri lihtsustatud, põhimõtteline skeem (Allikas: CO2meter.com)

Kokkuvõtlikult võib öelda, et NDIR andurite täpsusprobleemid on laiemalt teada ning nende eemaldamiseks on välja pakutud nii füüsilisi kui ka matemaatilisi kompensatsioonimeetmeid, mida osaliselt uuemates andurites osaliselt ka rakendatud on. Siiski ilmneb mitmetest CO₂ mõõteandmeid analüüsivatest uurimustöödest, sh käesoleva töö andmestikust, et varieeruvus andmete kvaliteedis on üldpildis suur.

3.5 EN 16798-1:2019 ja TAIL meetodika rakendamine õhukvaliteedi hindamiseks

EN 16798-1:2019 „Sisekeskkonna lähteandmeid energiatõhususe projekteerimiseks ja hindamiseks, lähtudes siseõhu kvaliteedist, soojuslikust keskkonnast, valgustusest ja akustikast“ on palju viidatud laialdaselt rakendust leidnud standard, milles esitatud meetodikat kasutatakse ühtlasi ka sisekliima jooksvaks hindamiseks pärast hoone kasutuselevõttu. EPBD juhiste ajendil on Wargocki et al. arendanud standardile tugineva tervikliku sisekliimakvaliteedi hindamismetoodika nimega TAIL, mille üheks osaks on ka CO₂-põhine õhukvaliteediklassi arvutamine .

Õhukvaliteediklass on võimalik arvutada igale ruumile või tsoonile, kasutades seda esindava CO₂ anduri logitud aegridasid valitud ajaperioodil. Mõõteandmed tohivad sisaldada ainult kasutusaegseid mõõteväärtuseid, seega tuleb kasutusajavälised perioodid mõõteandmestikust enne arvutust eemaldada.

Sisekliimaklasside, sh õhukvaliteediklasside kategooriaid on neli. Klassid on seotud asukate võimalike ootustega. Klassile I vastab kõrge ootuste tase, klassile II keskmine, klassile III tagasihoidlik ning klassile IV madal. Normaalseks tasemeks loetakse taset „keskmine“.

Klasside arvutamisel jaotatakse kasutusaegsed mõõteandmed nelja erinevasse ppm-väärtuste vahemikku. Vahemike piiride määratlemiseks on EN standardis pakutud välja kolm erinevat CO₂-põhist meetodit, millest mõõtmistepõhise jookva jälgimise puhul on asjakohasem tunnetuslikul õhu kvaliteedil põhinev meetod (meetod 1). Sama meetodit

kasutatakse ka LEVEL(s) raamistikus. Vahemikke eraldavad väärtused on esitatud teoreetilise ideaalse baastaseme suhtes, milleks on välisõhu CO₂ kontsentratsioon. EN standardis on esitatud välisõhu CO₂ kontsentratsiooniks lihtsustatult konstantne 400 ppm. TAIL metoodika märgib ära, et välisõhu CO₂ tase võiks olla ajas mõõdetud väärtus.

Tunnetusliku õhukvaliteedi meetodi järgselt on õhukvaliteedi klasside piirid järgmised:

- I. < baastase + 550 ppm
- II. baastase + 550 ppm... baastase + 800 ppm
- III. baastase + 800 ppm ... baastase + 1350 ppm
- IV. > baastase + 1350 ppm

Arvutuslik sisekliimaklass leitakse lihtsustatud TAIL metoodika järgi: kõrgem sisekliimaklass saavutatakse, kui sellest madalamates klassides ei asu summaarselt rohkem kui 5% kogu kasutusaegsetest andmetest. Originaalses TAIL metoodikas kehtis 5% taluvuspiir vaid vahetult madalamale klassile ning 1% taluvuspiir kaks klassi madalamale klassile.

Näide:

eeldame, et välisõhu CO₂-tase on konstantselt 420 ppm. Ühe nädala jooksul viibiti ruumis kokku 40 tundi. Mõõteväärtused ja nende suhtelised kestused jaotusid klassidesse järgnevalt:

Õhukvaliteediklass ja vastav ppm-vahemik	Mõõdetud tundide arv n vastavas ppm-vahemikus	Protsent kogu tundide arvust $\frac{n}{\Sigma n} \cdot 100\%$
I: 400...950	34	85%
II: 950...1200	4,4	11%
III: 1200...1750	1,4	3,5%
IV: 1750+	0,2	0,5%

Klassi I ei ole võimalik saavutada, kuna selles klassis asub 85% kogu perioodi mõõteväärtustest ning ülejäänud 15% asuvad madalamates klassides, mis ei mahu taluvuspiiri 5% sisse. Saavutatakse klass II, kuna klassis III ja IV asuvate mõõteandmete summaarne kestus jääb alla 5% piiri.

4. METOODIKA

Töö eesmärk on esitada statistiline taustinfosüsteem uuenevatele sisekliima hindamismetoodikatele ja -suunistele ning kasvavale CO₂-andurite kasutustrendile, et suurendada teadlikkust anduritega seotud põhiprobleemidest, nende mõjust mõõtetulemustele ning võimalikest kompensatsiooni- ja järeltöötlusmeetmetest.

Töö käigus hinnatakse mõõteandmete kättesaadavust ja kvaliteeti, tutvutakse iseloomulike käitumismustritega, pakutakse välja järeltöötusalgoritmid ja optimeeritakse nende tööd. Võrreldakse töödeldud ja töötlemata andmeid ning nende mõju tulemustele lähtuvalt lähtuvalt Euroopa standardi EN 16798-1:2019 hindamismetoodikast.

Statistilise üldistusvõime saavutamiseks on töösse kaasatud mõõteandmed võimalikult paljudelt erinevatelt anduritelt maksimaalse ajaperioodi jooksul. Eelnevat korrastamist vajab andmete struktuur ja vorming, kuid puudulike või vigaste andmete hulk ja mõju on osa uurimisobjektist. Andmetöötluseks kasutatakse programmeerimiskeele Python 3.10.4 laialtlevinud andmetöötlusmoodulit Pandas 2.0.3. Andmed visualiseeritakse kasutades Python'i mooduleid Matplotlib 3.7.1 ja Seaborn 0.12.2.

Esmalt hinnatakse andmete üldist kättesaadavust, pidevust ja lihtsasti tuvastatavate vigade tüüpe ning osakaalu. Etapi eesmärgiks on anda ettekujutus tüüpiliste andurite töökindlusest ja võimalikest andmeedastusprobleemidest EPBD monitoorimisnõuete valguses ning eemaldada selgelt vigased andmepunktid valimist.

Seejärel rakendatakse baastaseme korrektsioonialgoritmi (BLC algoritmi), et eemaldada andurite puudulikust kalibreerimisest, pikaajsest kõrvalekaldumisest vm mõjuteguritest tulenev nihe. Optimeeritakse algoritmi parameetreid, et saavutada eesmärgipärane ja suure usaldusnivooga töö võimalikult laia andmestiku ulatuses. Tulemusi võrreldakse nii mõõtühikute skaalal (ppm) kui ka sisekliimaklassidesse jaotatult (lähtuvalt LEVEL(s) ja EN 16798-1:2019 arvutusmetoodikatest).

Pärast andmete töötlemist BLC algoritmiga rakendatakse kasutusaegade tuvastuse algoritmi. Antud töö andmestiku põhjal arendatakse edasi algoritmi töö täpsust ja kiirust. Kuna puuduvad tõesed andmed kasutusaegade valideerimiseks, testitakse tulemuste tundlikkust tuvastustäpsusele, laiendades või kahandades tuvastatud kasutusperioode kuni ühe tunni võrra. Sarnaselt BLC mõjuanalüüsile hinnatakse tuvastustäpsuse mõju, kategoriseerides tulemused sisekliimaklassidesse ning esitades kestuskõveratel.

4.1 Mõõteandmete kogumisprotsessid

Töös kasutatakse NDIR-tüüpi CO₂ andurite mõõteandmeid. Andmed on kogutud aastate vahemikus 2021...2024 erinevatest Eesti haridusasutustest, sh lasteaedadest. Algses valimis oli kokku 212 erinevat andurit 17 hoonest andmestiku kogumahuga 333 aastat. Andurite jaotuvus on hoone tüüpide järgi esitatud tabelis 4.1. Tabelis olev struktuur säilib ka kõigil järgnevatel joonistel, kus on horisontaalteljele märgitud „212 erinevat andurit“, näiteks joonisel 5.1.

põhi- ja keskharidusasutused
lasteaiaid
kõrgharidusasutused

Tabel 4.1 Andurite ja hoonetüüpide ülevaade

hoone tüüp	pkh		la			pkh		la		kõh								
andurite arv hoones	57		5	17	7	5	5	10	6	15	1	6	14	1	7	11	21	24

Andurite paigutamine, andmete kogumine, märgistamine ja talletamine oli teostatud osana projekti DigiAudit arendustööst. Ruumide valim, andurite mudelite ja paigutuse valik ning andmeedastuslahendused on määratud eraldiseisvalt antud uurimistööst. Töö eesmärk on kasutada olemasolevaid lahendusi ja andmeid printsiiobil „on nagu on“, et anda ülevaade võimalikest eesootavatest probleemidest ja täheplepanu vajavatest aspektidest EPBD uute nõuete rakendamise valguses turutingimustel, mitte laboratoorselt kureeritud katsete näitel.

Andmed on kogutud andmepäringuna ühtsest andmebaasist (v.a. üks hoone, mille andmestiku edastas hoone esindaja isiklikult), kasutades „CO₂“ kategooriatunnusega andmeid maksimaalse saadaoleva ajaperioodi ulatuses.

4.2 Mõõteandmete puhastamine ning andmekadu

Enne algoritmide rakendamist ja sisekliimaklasside arvutamist tutvutakse andmetega, visualiseerides need

- 1) ajateljel
- 2) ppm-väärtuste esinemissageduste jaotustena
- 3) ajasammude esinemissageduste jaotustena

Tuginedes visualiseeringutelt kogutud vihjetele väljapaistvate anomaaliade kohta, kavandatakse edasine andmete puhastusmetoodika vajaduspõhiselt. Metoodika koos protsessi ning vahetulemustega on esitatud peatükis 5. Igas puhastusetapis eemaldatud või järelejäänud andmete protsent ning vastav analüüs on esitatud peatükis 6.1.

4.3 Baastaseme korrektsioon

Baastaseme korrektsioonialgoritmi (BLC algoritmi) eesmärk on ühtlustada mõõteandmete algoritmiliselt tuvastatud baasnivoo teoreetilise baasnivooga 400 ppm. Baasnivoo on siseruumi minimaalne teoreetiline võimalik CO₂ tase tulenevalt välisõhu CO₂ kontsentratsioonist. Algoritmi rakendamise vajadus tuleneb tõdemustest, et:

- standardijärgsed siseõhukvaliteedi klassid tuleb arvutada teoreetilise baasnivoo ehk välisõhu ppm-kontsentratsiooni suhtes;
- NDIR-tüüpi andurite käitumises on nii vanemas kui ka uuema teaduskirjanduses märgitud deklareeritud täpsusele mittevastavust, vigast kalibreeringut, anomaalseid käitumismustreid ning kalibreeringu kaldumist ajas. Seda kinnitavad ka antud töö raames uuritavad laiaulatislikud mõõteandmed.

4.3.1 Algoritmi tööpõhimõte

Baastaseme korrektsioonialgoritmi töö põhineb eeldusel, et liikuva akna ulatuses saavutatakse CO₂ baastase. Algoritmi sisendiks on 2 tööparameetrit ning mõõteandmete vektor:

$t_{win,h}$ - liikuva akna suurus, h

p_{low} - alumise protsentiili väärtus, %

$x(t)$ on mõõteandmete vektor, kus $t = 1, 2, \dots, N$

CO₂ mõõteandmete aegrea $x(t)$ ulatuses rakendatakse parempoolse joondusega liikuvat akent suurusega $t_{win,h}$, milles esinevate mõõteväärtuste alumise protsentiili p_{low} väärtus määratakse akna esmaseks tuvastatud baastasemeks $b_1(t)$:

$$b_1(t) = \text{protsentiil}(\{x(j)\}_{j=\max(1, t-t_{win,h}+1)}^t, p_{low})$$

Vihman et. al artiklis mainitud ajalise viibe vältimiseks nihutatakse $b_1(t)$ elemente ettepoole üks samm vähem kui akna suuruse võrra, tekitades vektori b_2 :

$$b_2(t) = b_1(t - t_{win,h} + 1)$$

Vektori b_2 puuduvad väärtused täidetakse sejärel tagasitäite meetodil (*backfill*), et täita võimalikud andmeaugud.

Baastaset defineeriva vektori $b(t)$ leidmiseks leitakse vektorite $b_1(t)$ ja $b_2(t)$ elementidevahelised maksimaalsed väärtused, ignoreerides võrdluspaarides esinevaid üksikuid puuduvaid väärtuseid:

$$b(t) = \max(b_1(t), b_2(t))$$

Sejärel arvutatakse korrektsiooniga mõõteandmete vektor $x'(t)$, lahutades originaalsest aegreast maha tuvastatud baastase $b(t)$ ning liites ideaalbaastase b_0 :

$$x'(t) = x(t) - b(t) + b_0$$

Olukorras, kus akna suurus moodustab üle 2/3 mõõteandmete vektori pikkusest, rakendatakse liikuva akna asemel ühekordne baastaseme arvutus kogu mõõteandmete vektori ulatuses, et vältida puuduvate väärtuste teket:

$$\text{kui } t_{win,h} > \frac{2}{3}N, \text{ siis } x' = \text{protsentiil}(x, p_{low})$$

Allpool on esitatud algoritmi kood programmeerimiskeeles Python. Parameeter $t_{win,h}$ on asendatud muutujaga 'window_size', mille ühikuks on sammud mõõteandmete järjendis 'ser'. Parameeter p_{low} on nimetatud kui 'low_perc' ning parameeter 'base' kajastab ideaalset baastaset. Parameeter 'min_periods' tähistab minimaalset nõutavat mõõteväärtuste hulka arvutusaknas, et arvutada protsentiilile vastav väärtus – nii välistatakse baastaseme arvutus suure andmekaoga perioodide puhul, kus usaldusnivoo oleks muidu äärmiselt madal.

```

1 import numpy as np
2 import pandas as pd
3
4 def fix_baseline(ser, window_size=288, low_perc=1, base=400):
5     """
6     :param ser: timeseries of measured co2 values, pd.Series
7     :param window_size: int, rolling window size in data steps, int
8     :param low_perc: int, low percentile value which is used as baseline value, int
9     :param base: ideal reference baseline value, int
10    """
11    len_ser = len(ser)
12    min_periods = int(window_size/2)
13
14    if window_size > int(len_ser*2/3):
15        return ser - ser.quantile(low_perc/100) + base
16    low_quant = ser.rolling(window_size, min_periods=min_periods).quantile(low_perc/100)
17    baseline = np.fmax(low_quant.values, low_quant.shift(-window_size+1).bfill().values)
18
19    return ser - baseline + base

```

4.3.2 Parameetrite optimeerimine ja algoritmi mõjuanalüüs

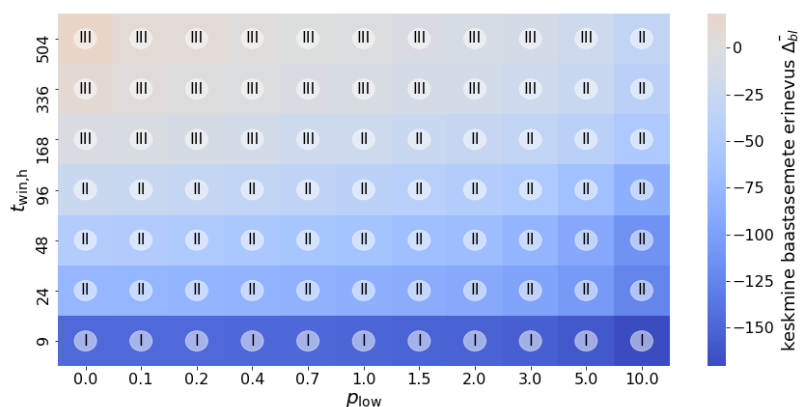
Allpool esitatakse BLC algoritmi mõjuanalüüsi optimeeritud tööprotsess. Tegelik protsess oli komplekssem, iteratiivsem ja sisaldas rohkemaid indikaatoreid. Algselt kavandatud protsess ei arvestanud äärmiselt varieeruva andmete karakteristikaga ja seetõttu tuli korrigeerida töö etappe, indikaatorite valikut ja parameetrite valiku optimeerimisülesannet.

Parameetrite optimeerimise ja mõjuanalüüsi tööprotsess

Parameetrite $t_{win,h}$ ja p_{low} optimeerimise ja mõjuanalüüsi lõplik tööprotsess on järgmine:

- 1) Kasutatakse andmeid stsenaariumist „01234567” (Joonis 6.1) – kõik andmepuhastusmeetodid ning kuni 30-minutiliste andmeaukude lineaarne interpoleerimine on rakendatud.

- 2) Kõigi andurite aegridasid töödeldakse BLC algoritmiga parameetrite kombinatsioonidel:
 - liikuva arvutusakna suurused (h): 9; 24; 48; 168; 336; 504
 - alumise protsentiili väärtused (%): 0; 0,1; 0,2; 0,4; 0,7; 1; 1,5; 2; 3; 5; 10
- 3) Määratakse primaarsed indikaatorid, mille järgi kvantifitseeritakse korrektsiooni mõju aegreale:
 - pideva skaalaga: keskmine baastasemete erinevus $\overline{\Delta_{bl}}$ (Valem 4.1)
 - kategoriseeriv: sisekliimaklass EVS-EN 16798-1:2019 järgi (I, II, III, IV)
- 4) Igale andurile arvutatakse indikaatorite väärtused punktis 2 esitatud parameetrite kõigi 66 kombinatsiooni puhul. Sisekliimaklass (tuvastatud kasutusaegadel) arvutatakse ka ilma korrektsioonita olukorrale. Väärtuste põhjal koostatakse anduri iseloomulikud maatriksid M_i ning $M_{i,ICC}$ (Joonis 4.1).
- 5) Igale andurile arvutatakse M_i põhjal sekundaarsete indikaatorite 3-elementiline vektor v_i .
- 6) Iga andur asetatakse 3D ruumi (kujutatuna kahel 2D hajuvusdiagrammil) teda iseloomustava sekundaarindikaatorite vektori v_i alusel. Hajuvusdiagrammidel eristatakse samasse hoonesse kuuluvaid andureid järgi värvikodeeringuga.
- 7) Analüüsitakse tekkinud jaotusmustreid, klastreid ja nende tekkepõhjuseid. Vaadeldakse ühekaupa hajuvusjaotusest väljavalitud andurite aegridasid koos vastavate iseloomulike maatriksitega. Analüüsitakse aegrea karakteristikat ja seost andurit iseloomustava maatriksiga ning paiknemisega hajuvusdiagrammil.
- 8) Teadvustatakse seosed andmete karakteristika ja parameetritele tundlikkuse vahel. Sellest tulenevalt püstitatakse optimeerimisülesanne, et anda soovitusel BLC algoritmi parameetrite valikuks üldistavas olukorras.
- 9) Illustreeritakse optimeeritud parameetritega algoritmi töö mõju ppm-kestuskõveratel ning ruumide standardijärgsete sisekliimaklasside jaotuste tulpdiaagrammidel.



Joonis 4.1 i -nda anduri iseloomulik maatriks M_i . Ringide sees on arvutuslikud sisekliimaklassid maatriksist $M_{i,ICC}$

Arvutusakna suuruste katseväärtuste valiku põhimõte

Arvutusakna $t_{win,h}$ katseväärtuste valikul lähtuti järgmistest põhimõtetest:

- 9 h – lühim eeldatav baastaseme taastumise aeg tüüpilise kasutusprofiiliga hoonetes, kus CO₂ taseme tuvastamine on oluline (kontorihooned ja haridusasutused, sh lasteaiad). 9 h on ka väljapakutud ajaaken Vihman et.al artiklis.
- 24 h – ööpäevane tsükkel. Päevase kasutusega hoonetes võiks öötundidest piisada, et kogu ruumiõhk ära vahetuks.
- 48 h – kaheööpäevane tsükkel (vaheväärtus).
- 96 h – neljaööpäevane tsükkel (vaheväärtus).
- 168 h – nädalane tsükkel. Maksimaalne eeldatav tsükli pikkus, mil baastase peaks igal juhul saavutatama tänu nädalavahetustele.
- 336 h – kahe nädalane tsükkel (vaheväärtus).
- 504 h – kolmenädalane tsükkel. Kolmekordne maksimaalse teoreetilise tsükli pikkus, et kontrollida tundlikkust eriti pikkade akende puhul.

Märkus: algne katseväärtuste valim piirdus 1 nädalaga (168 h), kuid valimit otsustati laiendada, tulenevalt esmaste tulemuste jaotusest väljapaistnud andurite aegridade karakteristika analüüsist.

Alumise protsentiili katseväärtuste valiku põhimõte

Alumise protsentiili p_{low} katseväärtuste valiku põhimõtted olid järgmised:

- 0% - võrdne arvutusaknas esineva minimaalse väärtusega.
- 1% - Vihman et. al väljapakutud väärtus.
- 10% - väärtus, mille sisse mahtusid mõningates aegridades esinenud äärmiselt spetsiifilised anomaalsed käitumismustrid.

Ülejäänud väärtused on vaheväärtused, kusjuures suurem resolutsioon madalamatel väärtustel on valitud tulenevalt seal esinevast suuremast tundlikkusest.

Primaarindikaatori 'keskmine baastasemete erinevus' arvutamine

Keskmine baastasemete erinevus $\overline{\Delta_{bl}}$ näitab ühe anduri i aegrea $x_i[t]$ keskmist elementidevahelist erinevust võrreldes aegrea BLC algoritmiga töödeldud versiooniga $x'_i[t]$. Kui muuta BLC algoritmi tööparameetreid, muutub ka keskmine baastasemete erinevus.

$$\overline{\Delta_{bl}} = \frac{1}{N} \sum_{t=1}^N (x'_i[t] - x_i[t]) \quad (\text{Valem 4.1})$$

kus $\overline{\Delta_{bl}}$ on keskmine baastasemete erinevus, ppm

$x_i[t]$ on i -nda anduri CO₂ mõõteandmete aegrida, ppm

$x'_i[t]$ on i -nda anduri BLC algoritmiga töödeldud aegrida, ppm

N on aegrea elementide arv

Autor on teadlik, et indikaatori puuduseks on positiivse ja negatiivse korrektsiooni vastastikune tühistav mõju aritmeetilise keskmise leidmisel. Algselt katsetati töö käigus ka mitmeid teisi indikaatoreid, sh keskmist baastaseme absoluutset erinevust, baastaseme hajuvust jt, kuid $\overline{\Delta_{bl}}$ osutus neist kõige paremini praktiliselt tõlgendatavaks, samuti aitas see kaasa optimeerimisülesande püstitamisele. Täiendavate indikaatorite kaasamine lisaks töösse märkimisväärselt komplekssus, kuid sisuline panus oleks minimaalne. Töö käigus otsiti aegridasid, kus võiks esineda ajas märki muutva korrektsiooni tühistav mõju. Selliseid aegridasid leidis, kuid vähe, ning isegi nende puhul ei esinenud täielikku tühistumist, vaid säilis nullist piisavalt erinev $\overline{\Delta_{bl}}$ väärtus, et indikeerida baastaseme kõrvalekaldeid teoreetilisest.

Anduri iseloomuliku maatriksi ja vektori koostamine

Anduri iseloomulik maatriks M_i on 6x11 maatriks, mis sisaldab indikaatori $\overline{\Delta_{bl}}$ väärtuseid kõigil parameetrite $t_{win,h}$ ja p_{low} etteantud kombinatsioonidel (kokku 66 kombinatsiooni). Maatriksiga M_i käib kaasas ka sarnane, sisekliimaklasse sisaldav maatriks $M_{i,ICC}$, mille väärtused on esitatud joonisel 4.1 ringides.

Sisekliimaklasside arvutamiseks on vaja tuvastada kasutusajad. Kuna kasutusaegade tuvastuse algoritmi töö sõltub BLC'st, on kõigile joonisel 4.1 esitatud stsenaariumitele rakendatud optimeeritud parameetritega (336 h ja 1%) BLC algoritm enne kasutusaegade tuvastamist. See tähendab, et sisekliimaklasside arvutamiseks vs kasutusaegade tuvastamiseks kasutatavate aegridade BLC-korrektsioonid on eraldiseisvad. Nii välistatakse kasutusaegade tuvastuse algoritmi sõltuvus BLC algoritmi parameetrite muutumisest, ehk sisekliimaklassid arvutatakse igas stsenaariumis täpselt samadel kasutusaegadel.

M_i järgi koostatakse anduri iseloomulik vektor v_i , mis sisaldab kolme sekundaarindikaatorit:

$$v_i = \begin{pmatrix} \overline{\Delta_{bl}} \\ \overline{p_{low}} \\ \overline{t_{win,h}} \end{pmatrix} \quad (\text{Valem 4.2})$$

kus:

aritmeetiline keskmine leitakse vastavalt:

$$\overline{\Delta_{bl}} = \frac{1}{66} \sum_{j=1}^6 \sum_{k=1}^{11} M_i(p_{low,k}, t_{win,h,j})$$

keskmise gradiendi x-komponent leitakse vastavalt:

$$\overline{p_{low}} = \frac{1}{60} \sum_{j=1}^6 \sum_{k=1}^{10} (M_i(p_{low,k+1}, t_{win,h,j}) - M_i(p_{low,k}, t_{win,h,j}))$$

keskmise gradiendi y-komponent leitakse vastavalt:

$$\overline{t_{win,h}} = \frac{1}{55} \sum_{j=1}^5 \sum_{k=1}^{11} (M_i(p_{low,k}, t_{win,h,j+1}) - M_i(p_{low,k}, t_{win,h,j}))$$

Hajuvusdiagrammide piirkondade eristamismetoodika

Joonistel 6.1 ja 6.2 avaldusid kaks huvipakkuvat eraldist: põhiklaster ja telgjoon. Piirkondade matemaatiliseks defineerimiseks kasutati vastavalt kahte meetodikat.

Põhiklaster defineeriti sekundaarindikaatorite $\overline{v_{t_{win,h}}}$ ja $\overline{v_{p_{low}}}$ tasandi järgi kasutades DBSCAN algoritmi, kusjuures enne algoritmi kasutamist $\overline{v_{t_{win,h}}}$ ja $\overline{v_{p_{low}}}$ jaotused standardiseeriti (keskmine väärtus 0 ja standardhälve 1). Klaster on piiritletud joonisel 6.2 halli katkendjoonega.

Telgjoon defineeriti sekundaarindikaatorite $\overline{v_{t_{win,h}}}$ ja $\overline{\Delta_{bl}}$ tasandil. Joone sobitamiseks kasutati iteratiivset lineaarse regressiooni meetodit, kus iga iteratsiooni (kokku 3) käigus jäeti välja 10% kõige madalama nõustuvusega anduritest. Viimasel iteratsioonil järelejäänud andurid, mis defineerivad lõpliku joone, on märgitud joonisel 6.4 punasega.

Optimeerimisülesande püstitamine

Püstitatakse nii kvantitatiivne kui ka kvalitatiivne optimeerimisülesanne. Mõlemad meetodikad on kaudsed, kuna puuduvad tõesed andmed otseseks valideerimiseks.

Kvantitatiivne optimeerimine

Kasutades töö lisas 1 esitatud meetodikat, sobitatakse CO₂ eksponentsiaalsete languskõverate järgi piirväärtused ehk tasakaalukontsentratsioonid, mis kajastavad tegelikku baastaseta. Otsitakse baastaseme korrektsioonialgoritmi parameetreid, mis minimeerivad viga tasakaalukontsentratsioonide järgi tuvastatud baastasemega.

Kvalitatiivne optimeerimine

Püstitatakse kvalitatiivne optimeerimisülesanne. Meetodikat aitab formuleerida käesoleva töö raames kogutud informatsioon, millele tuginedes valitakse kaks heuristilist meetodit: *Pareto* printsiip ning *Occam'i* parsimoonia. Mõlemad lähenemised on laialdaselt kasutatust leidnud nii elulistes, filosoofilistes kui ka teaduslikes käsitlustes, sealhulgas ka keerukamate andmetöötlus- ja masinõppeprobleemide lahendamisel. Optimeerimisel ei otsita optimumpunkti, vaid pragmaatilisi lahendusi, mis saavutavad ka piisava rahuldavuse kriteeriumi (*satisficing*). Kuna protsess on tihedalt seotud vahetulemustega, esitatakse see tulemuste sektsioonis alapeatükis 6.2.3

Andurite eristamine hoonete järgi

Ühe hoone raames on suurem tõenäosus sarnaste füüsiliste mõjutegurite olemasoluks: andurimudel- ja vanus, BMS, ventilatsioonisüsteemi tüüp ja töörežiim, tehnosüsteemide seisukord, hoone kasutusprofiil. Seega eristatakse andurite hoonesse kuuluvust sekundaarindikaatorite hajuvusdiagrammidel värvikodeeringu järgi, kuna see aitab leida seoseid andurite andmekarakteristika ja füüsiliste ühiste mõjutegurite vahel.

4.4 Kasutusaegade tuvastamine

Vastavalt EN 16798-1:2019 metoodikale arvutatakse sisekliimaklassid ainult kasutusaegsete mõõteandmete järgi, millest tuleneb vajadus hinnata või tuvastada ruumide kasutusaegasid. Lihtsaim lahendus oleks kasutada konstantset, tüüpilist kasutusprofiili, näiteks majandus- ja taristuministri määruse nr 58 „Hoonete energiatõhususe arvutamise metoodika“ §6 tabelis 1 esitatud kellaaegasid ja nädalapäevi vastavalt hoone või ruumi tüübile. Autor tõdes aga eelnevas uurimustöös, et kasutusaegade tuvastamise täpsus on suure määravusega sisekliimaklasside arvutamisel ning et eelistada tuleks dünaamilisi meetodeid.

Kasutusaegade tuvastamise meetodeid, sh CO₂-põhiseid algoritme on olemasolevas teaduskirjanduses välja pakutud mitmeid, kuid valdavalt on tegu võrdlemisi keerukate ja masinõppepõhiste meetoditega või on vajalik liikumisandurite kasutamine. Autor arendas varemtehtud uurimistöö raames välja uue, alternatiivse CO₂-põhise meetodi inspireerituna Pedersen et.al väljapakutud metoodikast, kuna olemasolev lahendus nõudis kõrget andmete salvestamise resolutsiooni, suurt arvutusmahukust ja tsoonipõhiselt kalibreeritud parameetreid. Käesoleva töö raames on algoritmi uuendatud: parandatud on üks valepositiivseid tuvastusi põhjustav viga ning märkimisväärselt on optimeeritud algoritmi töökiirust.

4.4.1 Algoritmi tööpõhimõte

- 1) Aegreale rakendatakse libiseva keskmise arvutus, et vähendada müra. Libiseva akna suuruseks on 30 minutit.
- 2) Arvutatakse aegrea muutused (diskreetne tuletis).
- 3) Muutuste aegrea igale hetkele arvutatakse tõusu- ja languslävendid, mis sõltuvad lineaarselt CO₂ absoluuttasemest etteantud muutumisvahemiku piires (Joonis 4.2).

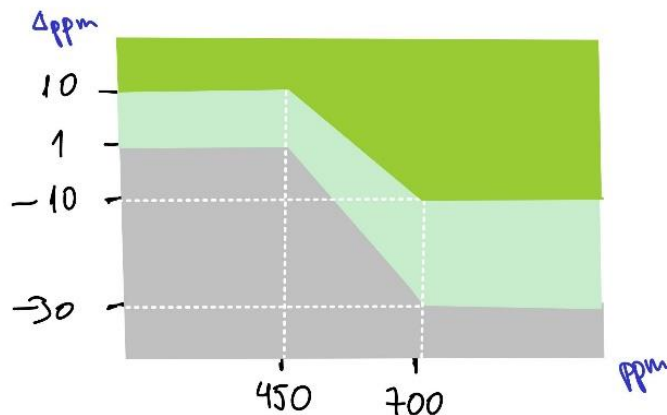
4) Muutuste aegrea elemente võrreldakse vastavalt tõusu- ja languslävendite aegrea elementidega.

- Kui muutuste aegrea element ületab samale ajahetkele vastavat tõusulävendit (ehk väärtus asub joonisel 4.2 ülemises, rohelises alas), kategoriseeritakse vastav ajahetk „tõusvasse” perioodi.
- Kui muutuste aegrea element ei ületa talle vastavat tõusulävendit, kuid on suurem kui languslävend (ehk väärtus asub joonisel 4.2 keskmises, helerohelises alas), kategoriseeritakse vastav andmepunkt „stagneerunud” perioodi.

5) Tõusvate ja stagnaerunud perioodide järgi määratakse kasutusajad.

- „Kasutuses” klassifitseering määratakse tuvastatud tõusuperioodidele ning neile vahetult järgnevale stagnaerunud perioodidele. Sisuliselt on tegu hüstereesiga.

Muutuvate lävendite kasutamine põhineb tõdemusel, et kõrgeenenud CO₂ tasemete puhul on tõenäolisem, et ruum on kasutuses ka väiksema, muutumatu või isegi veidi langeva CO₂ trajektooriga. Samas, madalamatel kontsentratsioonidel on vaja tugevamat tõusu, et indikeerida ruumide kasutust; analoogselt on vaja väiksemat langust, et väljuda stagnaerunud faasist. Lävendite väärtused ja muutumisvahemikud on kalibreeritud mitmekülgse karakteristikaga andmestiku põhjal ajaresolutsioonile 5...15 min. Hüsteresi kasutamine on inspireeritud Pedersen et. al metoodikast.



Joonis 4.1 Kasutusaegade tuvastusalgoritmi tõusu- ja languslävendite määratlemine ja hüsteresi, sõltuvalt CO₂ tasemest

Algoritmi edasiarendus

Käesoleva töö raames arendati olemasolevat algoritmi edasi.

- 1) Varasemalt oli arvutusaeg liiga pikk bilateraalse filtri kasutamise tõttu. Bilateraalne filter asendati lihtsa, lühikese libiseva keskmisega.
- 2) Kontsentratsioonide järgi muutuvad lävendid tekitasid valepositiivseid tuvastusi pikkadel langusperioodidel, kus CO₂ tase püsis kõrge. Sellised perioodid esinesid ruumides, kus ventilatsioon lülitus välja enne CO₂ taseme märkimisväärset alanemist ning äärmiselt aeglane taseme langus ei olnud kõrgetel kontsentratsioonidel piisav, et väljuda stagnaerunud perioodist. Probleemi lahendamiseks lisati algoritmi osa, mis tuvastab automaatselt sellised pikad

langusperioodid ning määrab need kasutusajavälisteks. Algoritm säilitab oma väikese arvutusmahukuse ja suure skaleeritavuse.

Algoritmi limitatsioonid

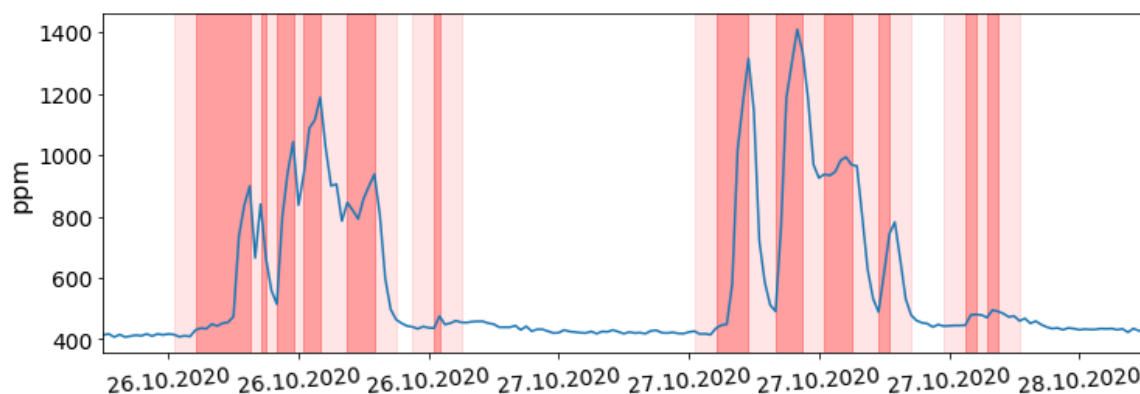
Algoritmi limitatsioonideks on

- valepositiivsed tuvastused: CO₂ aegridades võib esineda ebatavalisi mustreid, näiteks juhtudel, kui ventilatsioon lülitub välja liiga vara ja kontsentratsioon pole isegi langustrendi jõudnud. Selline olukord võib üksikutes järelejäänud olukordades endiselt põhjustada pikaajalise valepositiivse tuvastuse perioodi.
- valenegatiivsed tuvastused: madalatel kontsentratsioonidel on tuvastustäpsus mõnevõrra madalam ja kasutusajad võivad olla alahinnatud, kuna parameetrite optimeerimisel peeti olulisemaks valepositiivsete tuvastuste vältimist.
- sõltuvus ajasammust: parameetrite kalibreerimine sõltub andmete salvestusintervallist. Vajalik on eelnev *resampling* (ajasammu ümberarvutus) sobilikule resolutsioonile või parameetrite ümberkalibreerimine.
- vajadus rakendada eelnevalt BLC algoritmi: kuna kasutusaegade tuvastusalgoritmi töö sõltub andmete absoluutväärtustest, tuleb baastaseme nihe eemaldada.
- algoritmi töö tulemusi pole valideeritud tegelike kasutusaegade järgi.
- võib esineda mustreid aegridades, mida käesoleva töö andmestikus pole käsitletud. Kalibreeritud parameetrid ei pruugi sobituda väga ebastandardsete mustrite puhul kasutusaegade hindamiseks. Samas vajab selliste mustrite tuvastamine tugevaid taustateadmisi tegelikust kasutusprofiilist ning algoritmi komplekssus suureneks märgatavalt.
- algoritm ei suuda tuvastada vigast sisendinformatsiooni.

4.4.2 Tundlikkuse analüüs

Kasutusaegade tuvastuse algoritmi väljundi valideerimiseks puuduvad tõesed mõõteandmed ruumide kasutusaegadest. Algoritmi tööd hinnatakse läbi tundlikkusanalüüsi, kus laiendatakse või kitsendatakse kõiki tuvastatud kasutusperioode 15-minutiliste sammude kaupa kuni 60 minutini.

Joonisel 4.3 on esitatud näide, kuidas tuvastatud kasutusajad muutuvad 60-minutilise laiendamise puhul. Esialsed kasutusajad on märgitud tugeva punasega, laiendatud kasutusajad sisaldavad lisaks ka heledama punasega märgitud ajavahemikke.



Joonis 4.3 Tuvastatud kasutusaegade (tugev punane) ja nende 60-minutilise laienduse (helepunane) näide kahe päeva ulatuses.

Tulemuste võrdlemiseks koostatakse CO₂-mõõteandmete kestuskõverad erinevates kasutusaegade stsenaariumites:

- algoritmiliselt tuvastatud kasutusajad, sh laiendused ja kitsendused
- standardkasutusel põhinevad, staatilised kasutusajad:
 - kõik lasteaiad: kell 7...19, E-R
 - kõik ülejäänud haridusasutused: kell 8...16, E-R
 - v.a. vahemikus 15. juuni...14. august
- ilma kasutusaegade määramiseta

Graafikule märgitakse sisekliimaklasside piirid ning TAIL-metoodika põhine 5% klassiületuse taluvuspiir.

4.4.3 Mõju sisekliimaklassidesse jaotumisele

Kasutatakse kõiki tundlikkuse analüüsis esitatud stsenaariume. Igas stsenaariumis arvutatakse vastavalt EN-16798-1:2019 ja TAIL metoodikatele sisekliimaklass igale ruumile. Esitatakse ruumide sisekliimaklassidesse jaotumised kõigi stsenaariumite puhul.

4.5 Töö limitatsioonid

Kuigi töösse kaasatud mõõteandmete hulk on suur, võivad tulemuste kallutatust põhjustada järgmised valimi ühisfaktorid:

- kõik hooned paiknevad Eestis;
- hoonetüüpide valimis on vaid lasteaiad, põhi-, kesk- ja kõrgharidusasutused;
- kõik mõõteandmed peale ühe hoone on kogutud digitaalsesse andmebaasi läbi ühe teenusepakkuja.

Lisaks on raske eristada erinevate anomaaliade tekkepõhjuseid, kuna teave andurite paigutuse, mudelite, andmete kogumis- ja edastussüsteemide kohta on puudulik.

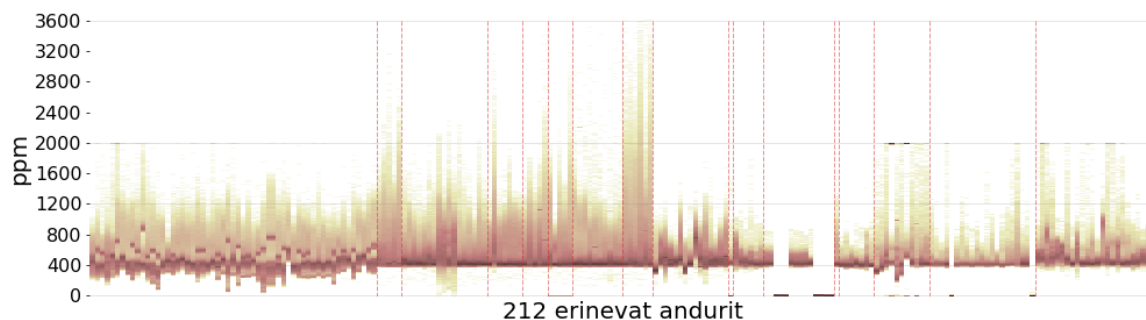
Väljapakutavate algoritmide töö ja lõpptulemuste valideerimiseks puuduvad tõesed andmed. Algoritmid on koostatud ja kalibreeritud parimate olemasolevate teadmiste järgi, et vähendada kriitilisema mõjuga vigade tekkeriski ja meetoodika kompleksust.

5. MÕÕTEANDMETE PUHASTAMINE NING ANDMEKADU

Enne algoritmide rakendamist ja sisekliimaklasside arvutamist tutvutakse andmetega. Esmased vihjed väljapaistvate probleemide ja erisuste kohta avaldusid, kui visualiseeriti andmeid

- 4) ajateljel
- 5) ppm-väärtuste esinemissageduste jaotustena (Joonis 5.1)
- 6) ajasammude esinemissageduste jaotustena (Joonis 5.3)

Ajateljel ja ajasammude jaotus kinnitasid, et andmed ei ole kaugeltki pidevad. Esines nii pikemaid kui ka lühemaid ajaauke. Samuti polnud kõik andmed salvestatud sama intervalliga. Jaotuste graafikult ja ajateljelt paistsid välja ka stagneerunud väärtused ning väärtused väljaspool skaalat. Nende probleemidega tegeletakse etapiliselt ning visualiseeritakse vastav andmekadu andurite lõikes. Iga järgneva puhastusstsenaariumi eelduseks on eelneva etapi läbimine.



Joonis 5.1 Töötlemata andmestiku ppm-väärtuste esinemissageduste jaotused eristatuna andurite lõikes. Vertikaalsed katkendjooned eraldavad hooneid. Värviskaala on logaritmiline.

5.1 Mõõteandmete puhastusetapid

5.1.1 Duplikaadid ja valemärgistused

Töö andmestikus esines suur hulk duplikaatandmeid ning mõningane hulk segadusttekitavalt tähistatud andmepunkte. Kõik sellised andmepunktid eemaldati valimist, kuid eemaldatud punktide hulka ei kvantifitseerita, kuna puudub statistiline üldistusvõime nende tekkepõhjuste kohta. Töö eesmärkidest lähtuvalt tuleb lihtsalt teadvustada, et suure hulga andmete haldamine, sh andmete märgistamine, kirjutamine ja lugemine andmebaasist võivad olla üheks komistuskohaks kvaliteetsete tulemusteni jõudmisel.

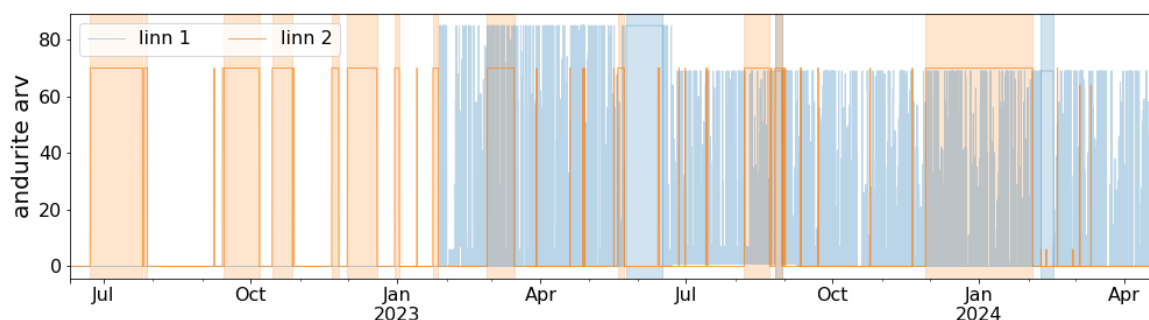
5.1.2 Mõõteandmete pidevus

Esmasel aegridade visualiseerimisel võis märgata mõningast anduritevahelist sünkroonsust pikemate andmeaukude puhul. Esimene andmeaukude sünkroonsuse

analüüs tehti hoonete lõikes, kuid sealt avaldus, et sünkroonsuse eristamiseks on hoonepõhine jaotus liiga detailne. Kuna hooned on jaotatud kahe Eesti linna vahel, rakendati kordusanalüüsil linnadesse grupeerimist. Avaldusid selged erinevused kahes linnas paiknevate hoonete andurite vahel (Joonis 5.2). Linnas nr 1 esinesid andmeaugud eranditult kogu süsteemi tasemel, linnas 2 esinesid andmeaugud andurite lõikes valdavalt asünkroonselt, kuid oli ka kaks sünkroonselt katkestust linna tasemel. Ühes hoones, mille andmed olid kogutud käsitsi, ei esinenud ühtegi andmeauku.

Töö andmestikust eemaldatakse linnapõhised sünkroonsed andmeaugud, mis on pikemad kui 24 h ning nende osakaal ei ole osa uurimisobjektist, kuna erinevate andmekogumislahenduste valim ei ole piisavalt suur, et teha üldistusi. Seega sisaldub pikkade andmeaukude eemaldamine joonisel 6.1 algstsenaariumis „012“.

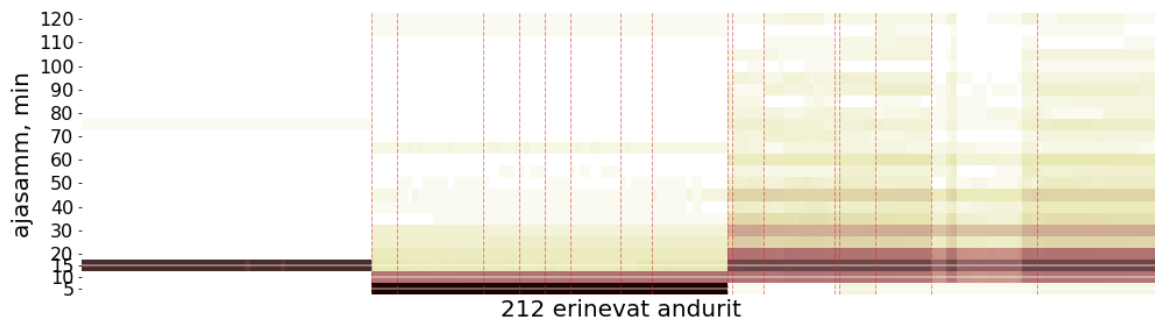
Andmeaugud, mis on lühemad kui 24 h, esitatakse osana uurimistööst kui „ajutised andmeedastusprobleemid“. 24 h piiri kriteerium valiti põhimõttel, et nii lühikesed andmeaugud peavad olema iselahenevad/ajutised, kuna <24 h on tõenäoliselt liiga lühike reageerimisaeg, et taastada süsteemide manuaalset sekkumist vajavad häiringud. Sellised ajutised andmeedastusprobleemid eemaldatakse stsenaariumi „0123“ raames (Joonis 6.1).



Joonis 5.2 Puuduvate andmetega andurite hulk ajahetkel, eristatuna kahe linna lõikes. Linnapõhised sünkroonsed pikad andmeaugud on markeeritud vastavalt värvitud perioodidena

5.1.3 Salvestamise intervall ja *resampling*

Joonisel 5.3 on kuvatud andurite mõõteandmete ajasammude esinemissagedused. Kaks peamist salvestusintervalli olid 5 ja 15 minutit, erinedes linnade vahel. Värviskaala intensiivsus on logaritmiline, et tuua paremini esile trendidest erinevate ajasammude jaotused. Ebastandardised ajasammud tulenevad peamiselt lühikestest andmeedastuse anomaaliatest. Pikkade ajasammude kaasamine graafikule ei ole antud peatüki huviobjektiks.



Joonis 5.3 Töötlemata andmestiku ajasammude esinemissageduste jaotused eristatuna andurite

Kogu andmestiku *resampling* ehk salvestusintervalli ümberarvutus teostatakse 15 minuti peale. Kõrgema sagedusega andmete puhul (5 ja 10 min) valitakse lähim esimene väärtus, kuna analüüs näitas, et võrreldes intervalli keskmistamisega säilivad andmete statistilised jaotuse näitajad (standardhälve, asümmeetriakordaja e *skewness* ja ekstsess e *kurtosis*) märgatavalt paremini. Keskmistamise puhul väheneks andmete dünaamika, sh asümmeetrilise saba ulatus jaotuse paremal poolel ehk kõrgemate ppm-väärtuste juures. Tippude ja järsema dünaamika säilimine on oluline kasutusaegade tuvastuse algoritmi töö ning sisekliimaklasside arvutamise seisukohast.

Resampling on joonisel 6.1 osana algstsenaariumist „012“.

5.1.4 Nullväärtuste eemaldamine

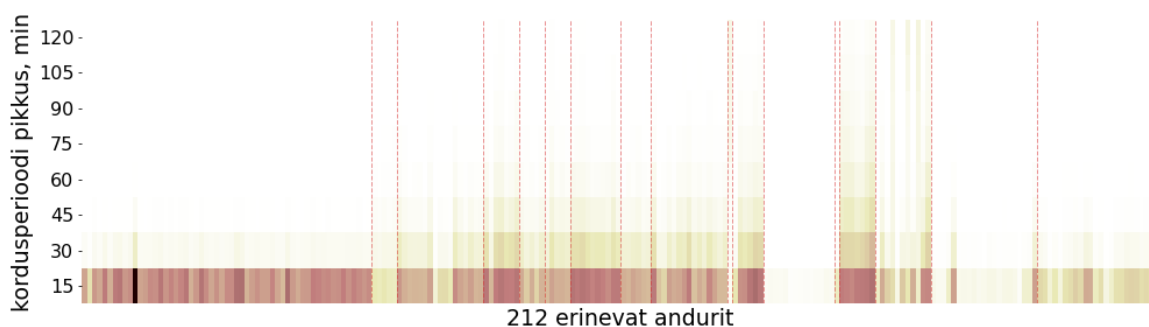
Andmete jaotustest joonisel 5.1 avadus mitmete andurite puhul intensiivsus 0 ppm juures, varieerudes peamiselt hoonepõhiselt. Lähemal vaatlemisel selgus, et mõned andurid edastasid aeg-ajalt täpselt 0-väärtusega andmeid. Sellised andmepunktid eemaldatakse stsenaariumi „01234“ raames (Joonis 6.1). Võimalik, et nullväärtuste olemasolu indikeerib kindlat anduritüüpi või hoones kasutatud andmete agregeerimissüsteemi.

5.1.5 Stagneerunud väärtuste eemaldamine

Järjekordne tüüpiline andmeedastusprobleem on stagneerunud ehk ajas muutumatud väärtused. Süsteem võib salvestada signaaliprobleemide esinemisel viimast teadaolevat mõõteväärtust. Selliste stagneerunud lõikude pikkuste jaotust pärast eelnevate puhastusmeetodite rakendamist näeb joonisel 5.4. Eemaldatakse stagneerunud väärtuste plokid, mille pikkus ületab 30 minutit. Teisisõnu, kui andur kordab täpselt sama mõõteväärtust rohkem kui 2 korda, loetakse muutumatud väärtused vigaseks plokiks ning korduvate väärtuste osa eemaldatakse. 30 minuti tolerants tuleneb tõdemusest, et vahepeal võib lugem anduri ppm-skaala resolutsioonist tulenevalt tõe poolest muutumatuna püsida. Eeldusel, et 30 minuti tolerants on valitud liiga lühike, võib siiski andmetöötlusega jätkata. On äärmiselt ebatõenäoline, et huvipakkuvatel,

kõrgenenud kontsentratsiooni järgi defineeritud kasutusperioodidel, mil ruumis esineb tugevaid CO₂ taseme kõikumisi, on mõõteväärtused 1 ppm täpsusega muutumatud. Lisaks on taastatakse kõik kuni 30 minuti pikkused andmeaugud stsenaariumis „01234567“, kasutades lineaarset interpoleerimist.

Stagneerunud väärtuste eemaldamise mõju järelejäänud andmehulgale on võimalik näha joonisel 6.1 stsenaariumis „012345“.



Joonis 5.4 Töötlemata andmestiku stagneerunud perioodide pikkuste esinemissagedused eristatuna andurite lõikes. Vertikaalsed katkendjooned eraldavad hooneid. Värviskaala on lineaarne

5.1.6 Skaalast väljasolevate väärtuste eemaldamine

CO₂ ppm-skaala minimaalne väärtus on teoreetiliselt välisõhu CO₂ tase ehk ca 380...440 ppm, vastavalt biomassi ööpäevasele ja aastaringsele süsiniktsükli. Siiski tuleks aktsepteeritavate väärtuste vahemiku alumine piir võtta oluliselt madalam, arvestades, et andurite baastase on sageli allapoole nihkunud. Kogu andmestiku jaotust detailselt uurides selgus, et nullväärtuse ümber esineb suurenenud kontsentratsioon väärtuseid, mis on tõenäoliselt vale tähisega salvestatud temperatuuriandmed. Minimaalse esinemissageduse punkt nende kahe jaotuse vahel asus ca 40 ppm juures, seega valiti 40 ppm ka skaala alumiseks piiriks. Skaala ülemiseks piiriks valiti sarnasel põhimõttel 3500 ppm, millest kõrgemate väärtuste esinemissagedus ei kujutanud enam kahanevat trendi, mis ühtiks ülejäänud jaotuskõveraga, vaid oli juhuslik. Sellest tulenevalt võib eeldada, et kõrgematel väärtustel puudub sidusus ülejäänud mõõteandmete dünaamikaga ja tõenäoliselt on tegu juhuslike vigadega.

Skaalast väljasolevate väärtuste eemaldamine toimub joonisel 6.1 etapis „0123456“.

5.1.7 Lühikeste andmeaukude täitmine

Etapis „01234567“ täidetakse lineaarse interpoleerimise abil kõik andmeaugud, mille pikkus ei ületa 30 minutit. Sellised andurid eristuvad joonistel, kus horisontaalteljele on märgitud „212 erinevat andurit“, sh joonisel 6.1, paremal pool. Etapi eesmärk on demonstreerida lihtsa ja suhteliselt kõrge usaldusnivooga meetodika rakendamist, et taastada suur osa andmetest, kus esineb pidevalt lühikesteid andmeaukude täitmist.

Lühikeste andmeaukude täitmine aitab parandada aegridade pidevust ja seeläbi ka järeltöötusalgoritmide töökindlust.

5.1.8 Järelejäänud anomaaliad

CO₂ kontsentratsioonide jaotustest joonisel 6.1 eristuvad intensiivsemad piirkonnad, mis ei langenud välja ka stagneerunud väärtuste eemaldamis etapis. Vastavate andurite aegridasid lähemalt uurides selgus, et andurid on loginud ligikaudu konstantset väärtust. Tõenäoliselt on tegu analoogset, mitte digitaalset päritolu veaga. Selliste vigade süstemaatiline tuvastamine on nüansirohkem ning nende olemasolu kuulub edasistes etappides töö uurimisobjekti hulka. Sellest tulenevalt ei eemaldata nimetatud anomaalseid perioode andmestikust. Perioodide mõju avaldub peatükis 6.2.1 esitatud hajuvusdiagrammidel ja teistel joonistel, kus eristatakse andureid põhiklastrisse kuuluvuse järgi.

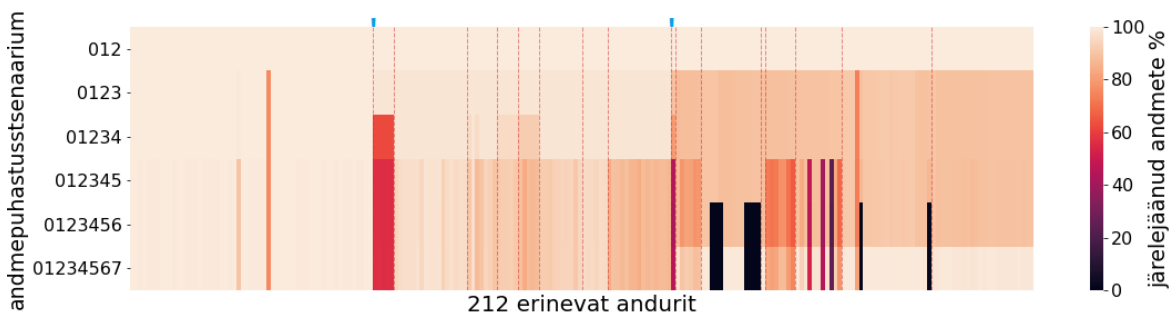
6. TULEMUSED JA ANALÜÜS

6.1 Mõõteandmete puhastamine ning andmekadu

6.1.1 Andmekao kvantifitseerimine

Eelnevas peatükis kirjeldatud andmete puhastustappide kumulatiivne mõju on kujutatud joonisel 6.1 kui järelejäänud andmete protsent iga anduri puhul, võrreldes algstsenaariumiga „012“. Andmepuhastustappide lühikirjeldused on järgmised:

- „012“ – eemaldatud on >24 h sünkroonsed katkestused andmeedastuses. Andmed on ümberarvutatud 15 min intervallile.
- „3“ – eemaldatakse <24 h juhuslikud katkestused andmeedastuses.
- „4“ – eemaldatakse nullväärtused.
- „5“ – eemaldatakse >30 min stagneerunud väärtused.
- „6“ – eemaldatakse skaalast 40...3500 ppm väljas olevad väärtused.
- „7“ – täidetakse kuni 30 min pikkused andmeaugud kasutades lineaarset interpoleerimist.



Joonis 6.1 Erinevates andmepuhastustappides tekkinud andmekao % andurite lõikes. Vertikaalsed katkendjooned eraldavad hooneid. Sinised markerid eristavad makrosüsteeme (linnad, andmekogumise meetoodika)

6.1.2 Analüüs

Stsenaariumi „3“ rakendamisel eristuvad siniste markeritega märgitud hoonete plokid. Plokke eristavad erinevad kõrgema taseme andmekogumismetoodikad: vasakpoolses plokis on andmed kogutud otse, ilma hoonest väljuva võrguülese edastuseta. Keskmist ja parempoolset plokki eristab paiknemine eri linnades, täpsem erinevus pole teada. Selgub, et lühikeste, alla 24 h pikkuste andmesalvestuse katkestuste olemasolu on tugevalt seotud kõrgema süsteemi tasemel valitud andmete koondamise meetoodikaga. Samuti on märgata, et manuaalselt kogutud andmetega hoonetes ei esine andmekatkestusi, mis viitab võrguühenduse probleemide mõjule andmete edastamisel.

Stsenaariumi „4“ rakendamisel, kus eemaldatakse nullväärtused, eristuvad selgelt üksikud hooned. See viitab hoone tasemel esinevatele ühisteguritele, nagu näiteks samad andurimudelid, ühine kohalik võrk või BMS.

Stsenaariumi „5“ rakendamisel, kus eemaldatakse stagneerunud väärtused, esineb samuti mõningane hoonepõhisus, kuid on ka üksikuid andureid, mis eristuvad hoone trendist. Võib järeldada, et stagneerunud väärtused tulenevad suurema tõenäosusega anduri või kohaliku andmeedastuse eripäradest. Stagneerunud väärtuste edastamine võib olla märk kohaliku võrgu ülekoormatusest või muudest häiringutest

Stsenaariumis „6“ eemaldatud CO₂-skaalast väljas olevate väärtuste olemasolu näitab, et mõned andurid olid valesti liigitatud ning edastasid tõenäoliselt hoopis temperatuuriandmeid.

6.2 Baastaseme korrektsioon

6.2.1 Hajuvusdiagrammidega tutvumine

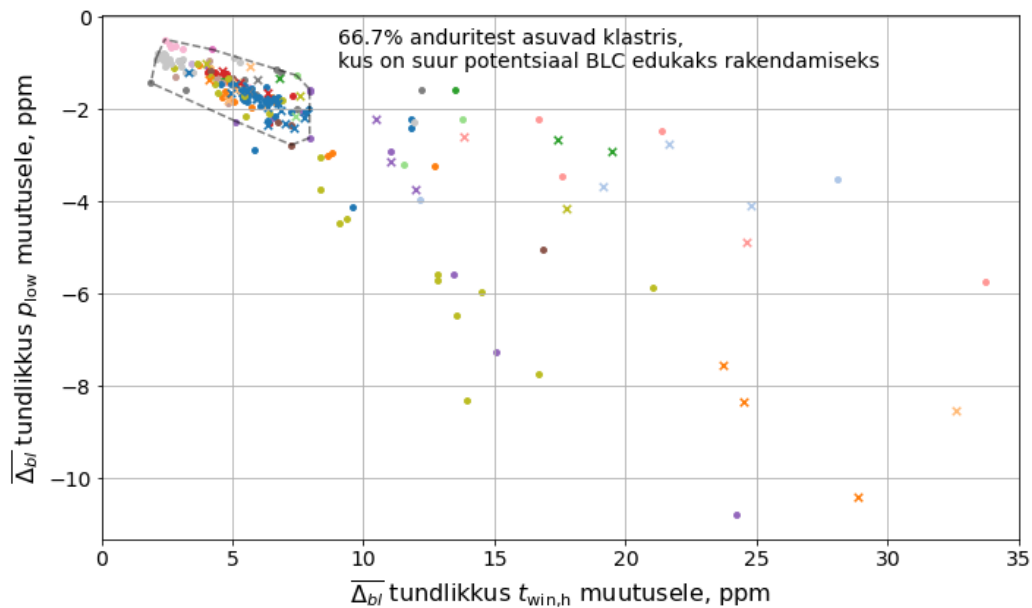
Valemi 4.2 järgi arvutati igale tsoonile 3-elementiline iseloomulik vektor v_i , mis kirjeldab baastaseme keskmise korrektsiooni $\overline{\Delta_{bl}}$ väärtuse tundlikkust algoritmi parameetrite varieerimisele.

Kuna sekundaarseid indikaatoreid on 3, esitatakse tulemused kahel eraldiseisval 2D hajuvusdiagrammil (Joonis 6.2 ja Joonis 6.3). Horisontaaltelg on mõlemal diagrammil ühine, seega võib neid kahte diagrammi kujutleda kui sekundaarindikaatorite 3D-ruumi eest- ja pealtvaateid.

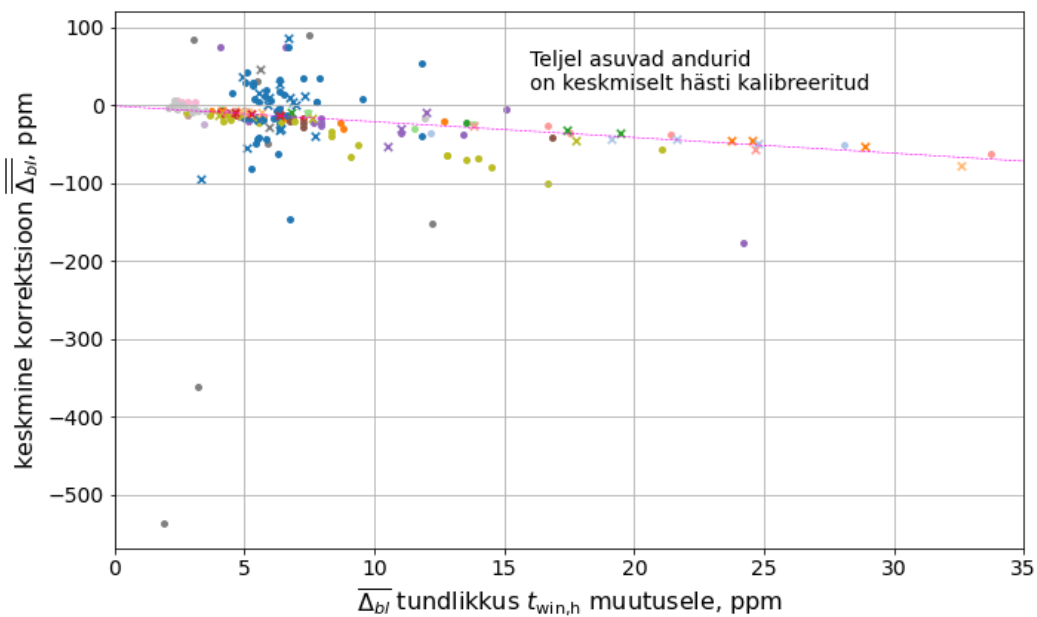
Iga punkt tähistab ühte tsooni, punkti värv tähistab kuulumist teatud hoonesse. Iga punkti asukoht põhineb vastaval iseloomulikul vektoril v_i . Lisaks on eristatud „X“ märgisega andurid, mille puhul muutus arvutuslik sisekliimaklass vähemalt ühe BLC algoritmi parameetrite kombinatsiooni korral (teisisõnu, kui maatriksis $M_{i,ICC}$ kujutatuna joonisel 4.1 ringide sees esines rohkem kui 1 erinev sisekliimaklassi väärtus).

Olgu joonis 6.2 edaspidi viidatud kui „eestvaade“ ja joonis 6.3 kui „pealtvaade“.

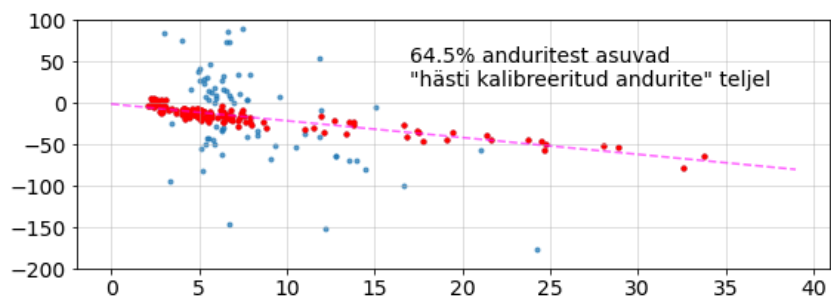
Kõikide telgede ühikud on ppm-skaalal. Kuna $t_{win,h}$ ja p_{low} arvutusparameetrid on valitud empiiriliselt, on esitatavad numbrilised väärtused vaid indikatiivsed ja ei oma füüsikaliselt tõlgendatavat vastet.



Joonis 6.2 Andurite jaotumine sekundaarindikaatorite hajuvusdiagrammil, „eestvaade“. Hooned on eristatud värvidega



Joonis 6.3 Andurite jaotumine sekundaarindikaatorite hajuvusdiagrammil, „pealtvaade“. Hooned on eristatud värvidega



Joonis 6.4 Hästi kalibreeritud andurite telge (roosa katkendjoon) defineerivad, ehk teljel asuvad andurid (märgitud punasega)

Eestvaate ja pealtvaate horisontaalteg

- Muutuja: anduri iseloomuliku maatriksi M_i gradient $\overline{\nabla t_{win,h}}$ liikuva akna suuruse $t_{win,h}$ varieerimisel.
- Matemaatiline tõlgendus: suurem $\overline{\nabla t_{win,h}}$ väärtus näitab, et andur on tundlikum liikuva akna suuruse $t_{win,h}$ varieerimisele. Positiivsus näitab, et akna suurendamisel suureneb ülespoole korrektsioon ja/või väheneb allapoole korrektsioon.
- Praktiline tõlgendus: korrektsioonialgoritmi eeldus on, et arvutusakna sees jõuab CO₂-tase tegelikule baastasemele ehk välisõhu CO₂-kontsentratsioonile. CO₂-taseme alanemine baastasemele toimub kasutusvälisel ajal. Taseme alanemise kiirus oleneb kasutusvälisel ajal esinevast õhuvahetusest. Suur tundlikkus arvutusakna suurusele võib viidata liialt vara väljalülituvale ventilatsiooniagregaadile, kuna infiltratsioonist tulenev õhuvahetus ei pruugi olla piisav baastaseme saavutamiseks enne ruumi uut kasutustsükli.

Eestvaate vertikaalteg

- Muutuja: anduri iseloomuliku maatriksi gradient $\overline{\nabla p_{low}}$ liikuva akna alumise protsentiili p_{low} varieerimisel.
- Matemaatiline tõlgendus: suurem absoluutväärtus näitab, et andur on tundlikum alumise protsentiili p_{low} varieerimisele. Negatiivsus näitab, et protsentiili suurendamisel suureneb allapoole korrektsioon ja/või väheneb ülespoole korrektsioon.
- Praktiline tõlgendus: korrektselt valitud arvutusakna suuruse, korrektselt funktsioneerivate ventilatsioonisüsteemide ja andurite puhul peaks alumise protsentiili mõju olema lähteülesandes valitud väärtuste vahemikus madal. Tundlikkus tekib, kui arvutusakna ulatuses esineb madalate väärtuste juures kõikumisi ehk madalate väärtuste koondumus võimaliku baastaseme ümber on kesine.

Pealtvaate vertikaalteg

- Muutuja: anduri iseloomuliku maatriksi keskmine väärtus $\overline{\Delta_{bl}}$ kõigil parameetrite $t_{win,h}$ ja p_{low} kombinatsioonidel.
- Matemaatiline tõlgendus: näitab kõigi parameetrite $t_{win,h}$ ja p_{low} kombinatsioonide puhul tuvastatud baastaseme keskmist kaugust ideaalbaastasemest (400 ppm).
- Praktiline tõlgendus: iseloomustab anduri keksmist kalibreeringut. Positiivsemad väärtused viitavad alakalibreeritud anduritele (korrektsioon toimub valdavalt ülespoole) ja negatiivsed väärtused viitavad ülekalibreeritud anduritele (korrektsioon toimub valdavalt allapoole).

Eestvaates avaldub ligikaudne lineaarne seos parameetrite $t_{win,h}$ ja p_{low} tundlikkuse vahel. Seos on oodatav, kuna kõrgem tundlikkus $t_{win,h}$ suhtes tähendab ühtlasi, et baastaseme leidmine ajas on keerulisem ning väärtused võimaliku baastaseme ümbruses on hajusamad, mis omakorda suurendab tundlikkust protsentiili p_{low} valikule. Siiski, ligi 2/3 kõigist anduritest koonduvad ühte klastrisse (joonisel piiritletud halli katkendliku joonega). Klastrisse kuulumine ei tähenda korrektset baastaset, vaid suurt potentsiaali selle korrektseks tuvastamiseks.

Pealtvaates võib märgata lineaarset trendijoont (roosa katkendlik joon). Kuna tsoonide koonduvus trendijoone ümber on selgelt eristuv, peab joone tekkemeetod olema tõenäoliselt numbriline, mitte statistiline. Joont defineerivad, sh ka „teljel asuvad“ andurid on märgitud joonisel 6.4 punasega.

Analüüsidel teljel paiknevaid ja teljest eraldiseisvaid andureid, selgus, et kõik teljel paiknevad andurid on keskmiselt hästi kalibreeritud (Joonis 6.5 ja Joonis 6.7). Tundlikkus arvutusakna suuruse suhtes tekib põhiliselt kasutusajavälise õhuvahetuskordsuse järgi. Samas, tihedamast punktipilvest eemaldunud ehk äärmiselt suure tundlikkusega andurite aegread viitavad tüüpilisele kasutusele mitteiseloomulikele mustritele CO₂-taseme dünaamikas või suuremale nüansirohkusele ventilatsioonisüsteemi töörežiimi(de) ja ruumi kasutuse vahel (Joonis 6.7).

6.2.2 Hajuvusdiagrammide piirkondade analüüs

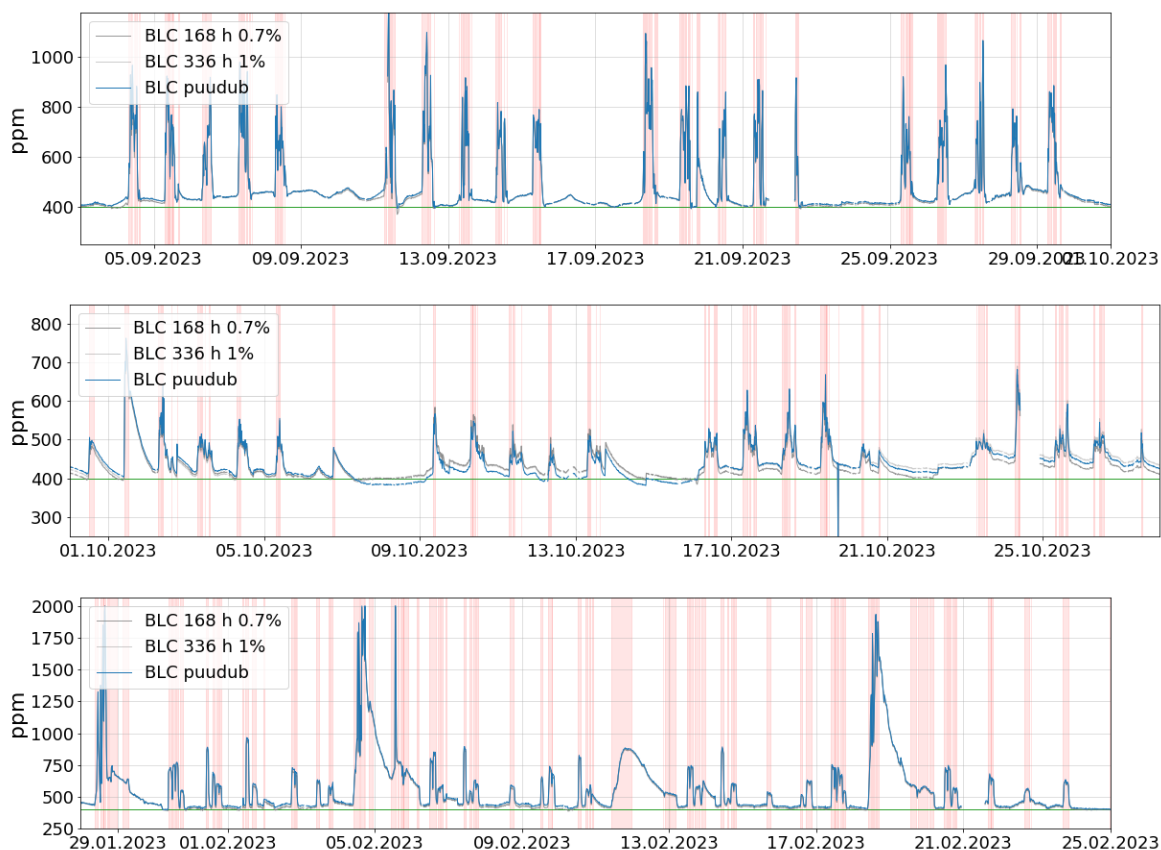
Hajuvusdiagrammide piirkondade tõlgendused on valideeritud, uurides visuaalselt paljusid aegridasid erinevatest diagrammidel eristatavatest stsenaariumitest:

- 1) andurid, mis asuvad klastris ja teljel;
- 2) andurid, mis asuvad klastris, kuid teljelt väljas;
- 3) andurid, mis asuvad teljel, kuid klastrist väljas;
- 4) andurid, mis ei asu teljel ega klastris.

Kõigile stsenaariumitele esitatakse illustratiivselt mõned iseloomulikud aegridade graafikud. Iseloomulikel graafikutel on valitud 4-nädalane ajaperiood, mis esindab kogu aegrea ulatuses esinevaid probleeme ja karakteristikat. Igale graafikule on halltoonidega lisatud ka kaks joont, mis illustreerivad BLC algoritmiga töödeldud aegrida vastavalt peatüki 6.2.2 järgi optimeeritud parameetritel (168 h 0,7%) ja (336 h 1%). Roheline horisontaaljoon kujutab ideaalse baastaseme referentsväärtust 400 ppm, mille suhtes on standardi EN 16798-1:2019 järgi defineeritud ka sisekliimaklassid. Punaselt markeeritud perioodid tähistavad algoritmiliselt tuvastatud kasutusaegasid peatükist 6.3. „Klastrid“ ja „teljed“ määratluse järgi koostatud andmete histogrammid on esitatud peatükis 6.2.3 alapeatükis „Histogrammid“.

Ruumid, mis asuvad klastris ja teljel

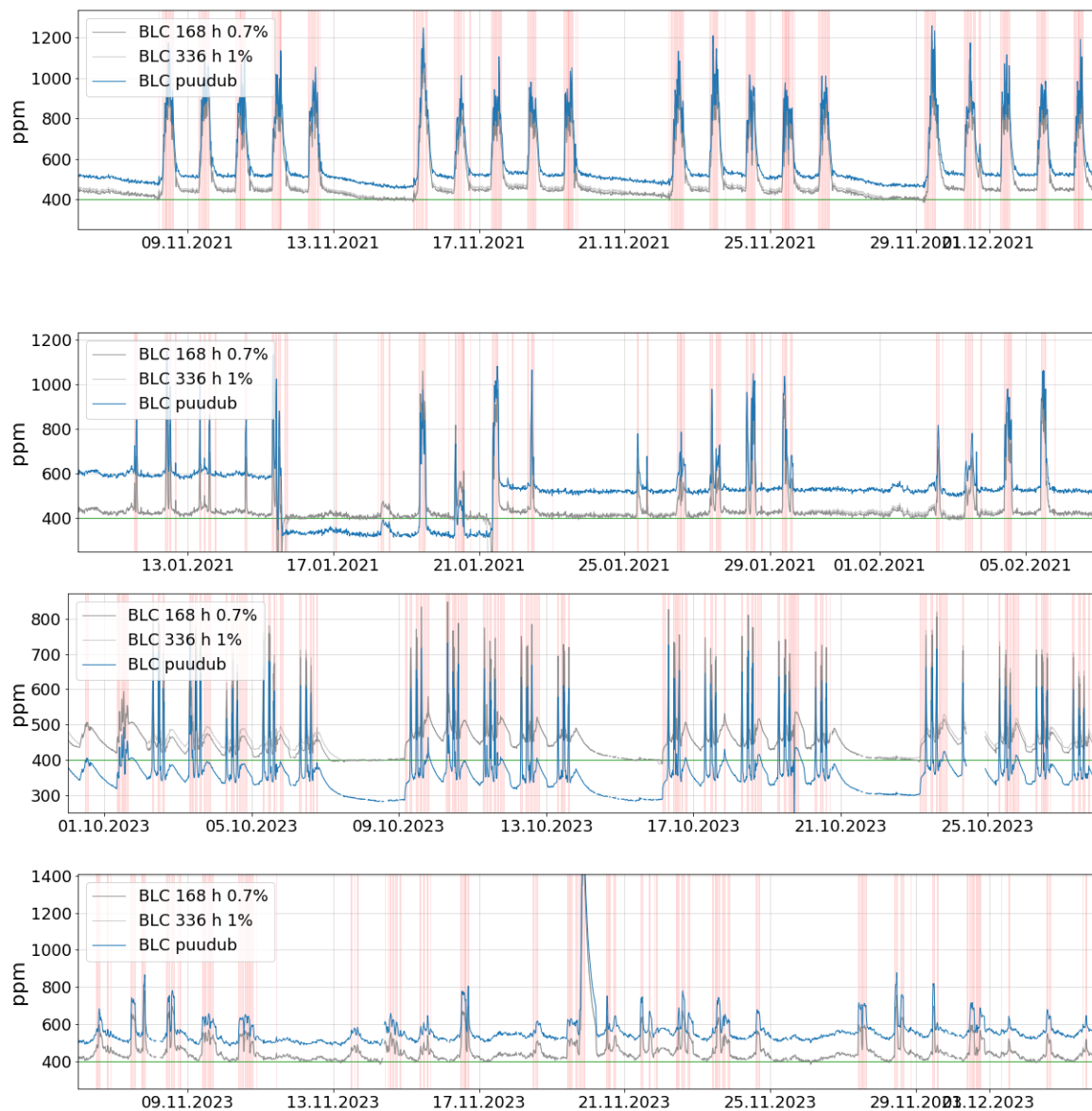
Ruumides, mille andurid paiknevad põhiklastris ja teljel, on suure tõenäosusega ootuspäraselt funktsioneeriv ventilatsioon, andmete karakteristikas ei esine suuri ebastabiilsusi ning andurid on keskmiselt hästi kalibreeritud. On kõrge potentsiaal edukaks BLC rakendamiseks ja kasutusaegade tuvastamiseks. Samas, BLC mõju on tõenäoliselt minimaalne.



Joonis 6.5 Aegride näidised: andurid, mis asuvad klastris ja teljel. Algoritmiliselt tuvastatud kasutusajad on märgitud punasega

Ruumid, mis asuvad klastris, kuid teljelt väljas

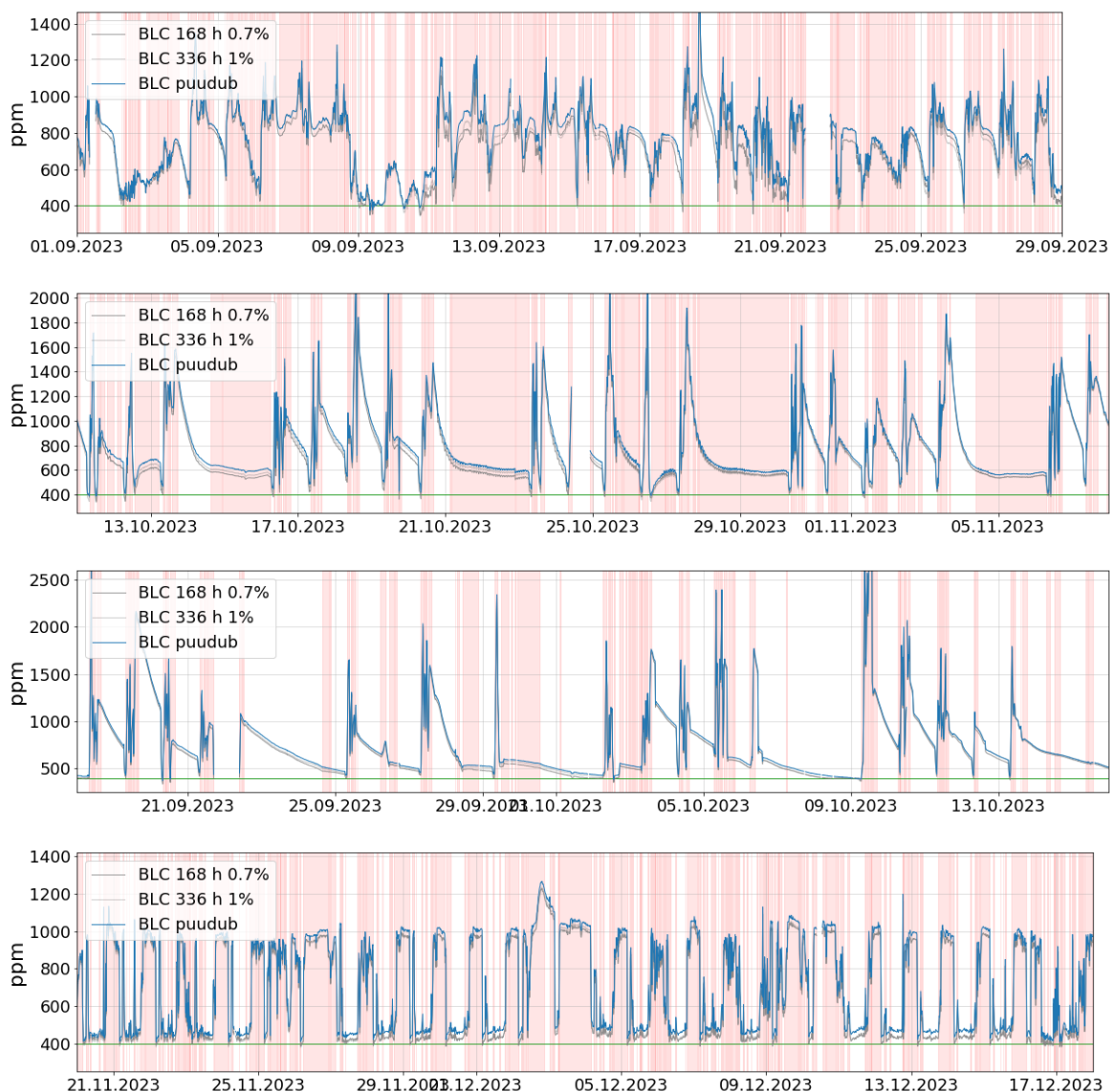
Ruumides, mille andurid paiknevad põhiklastris, kuid teljelt väljas, on samuti suure tõenäosusega ootuspäraselt funktsioneeriv ventilatsioon, andmete karakteristikas ei esine suuri ebastabiilsusi, kuid anduri kalibreering ei ole adekvaatne. On kõrge potentsiaal edukaks BLC rakendamiseks ja kasutusaegade tuvastamiseks. BLC kasutamine on väga vajalik.



Joonis 6.6 Aegridade näidised: andurid, mis asuvad klastris, kuid teljelt väljas. Algoritmiliselt tuvastatud kasutusaegad on märgitud punasega

Ruumid, mis asuvad teljel, kuid klastrist väljas

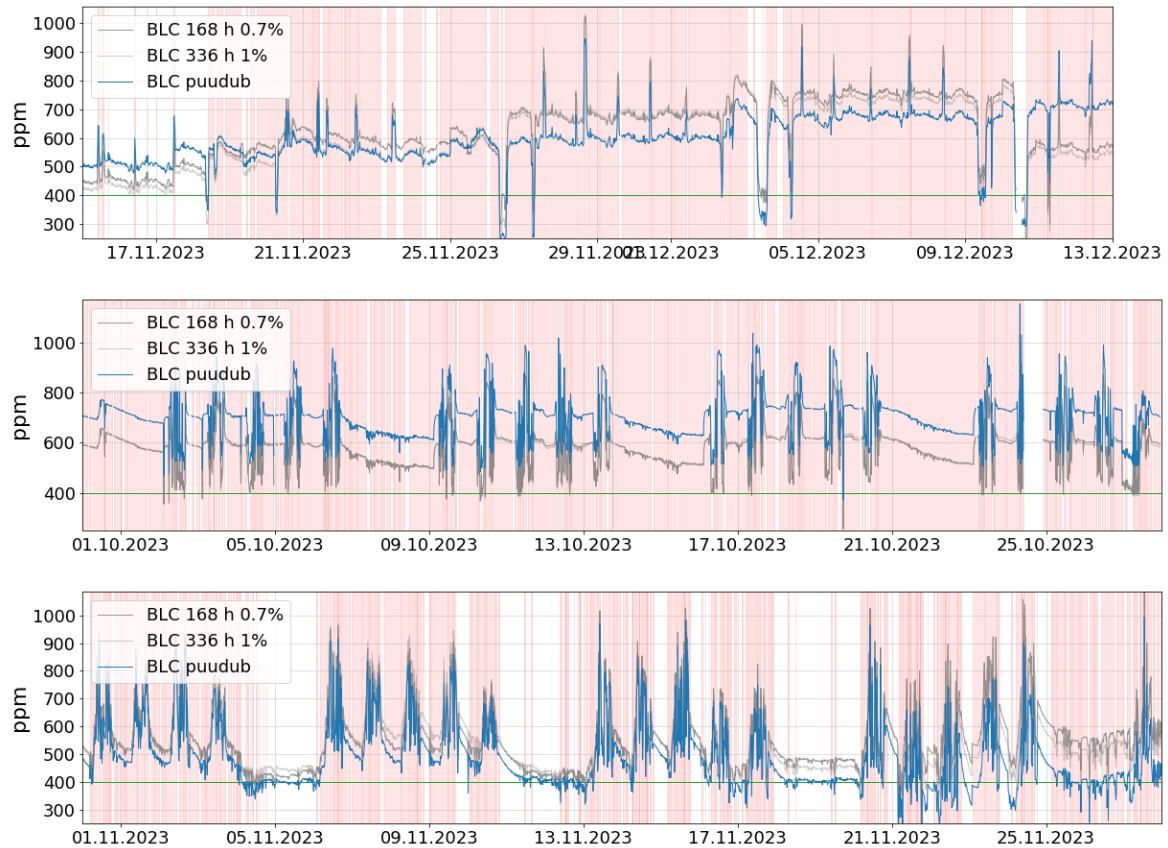
Ruumides, mille andurid paiknevad teljel, kuid põhiklastrist väljas, esinevad mustrid, mis viitavad suuremale nüansirohkusele ventilatsioonisüsteemi töörežiimi(de) ja ruumi kasutuse vahel. Andmete karakteristikas esineb komplekssemaid mustreid, mille tõlgendamine on keerulisem ning suurte mõõtevigade olemasolu pole välistatud. On keskmisest madalam potentsiaal edukaks BLC rakendamiseks ja kasutusaegade tuvastamiseks. Samas, BLC mõju on peatükis 6.2.2 leitud optimeeritud parameetrite puhul tõenäoliselt madal.



Joonis 6.7 Aegridade näidised: andurid, mis asuvad teljel, kuid klastrist väljas. Algoritmiliselt tuvastatud kasutusaegad on märgitud punasega

Ruumid, mis ei asu teljel ega klastris

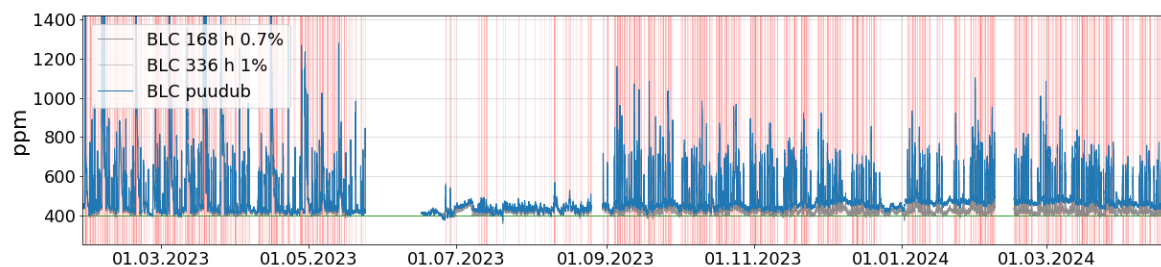
Nendes ruumides on olemasolevate andmete põhjal raske hinnata tegelikku olukorda. Andmete karakteristikas esineb märkimisväärseid ebastabiilsusi, mis viitavad probleemidele anduriga. On madal potentsiaal olemasolevate BLC ja kasutusaegade tuvastamise algoritmide edukaks rakendamiseks. Selliste andurite väljund ei ole hästi rakendatav praktiliste eesmärkide saavutamiseks, sealhulgas DCV juhtimiseks.



Joonis 6.8 Aegridade näidised: andurid, mis ei asu klastris ega teljel. Algoritmiliselt tuvastatud kasutusajad on märgitud punasega

Pikaaegne kaldumine

Jooniselt 6.9 on näha ühe anduri baasnivoo nihkumist pikema ajaperioodi jooksul. BLC algoritm sobib hästi probleemi lahendamiseks. Antud andur asus telje lähedal ning klastris.

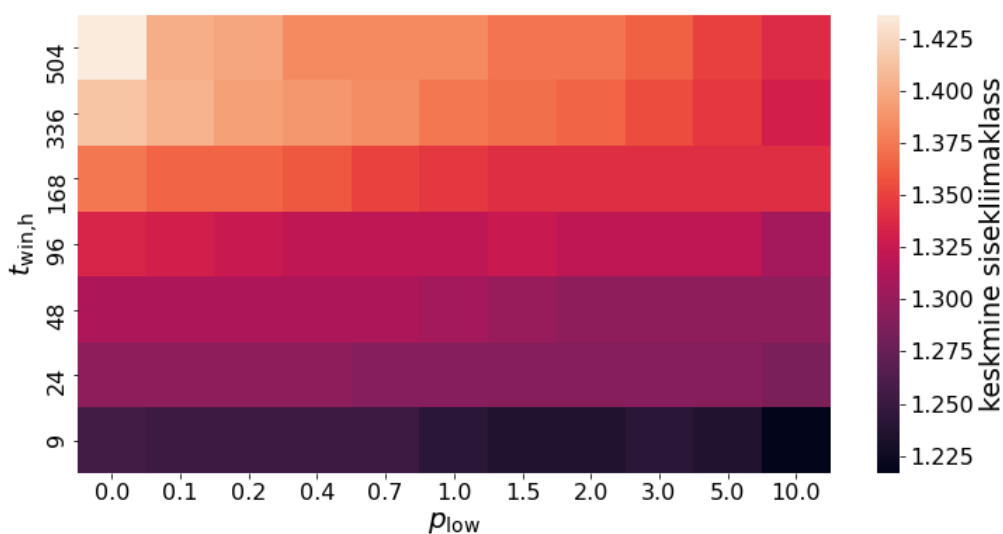


Joonis 6.9 Aegridade näidis: baasnivoo pikaaegne nihkumine. Algoritmiliselt tuvastatud kasutusajad on märgitud punasega

Mõju sisekliimaklassidele

Joonistel 6.2 ja 6.3 on märgitud tähisega „X” ruumid, kus arvutuslik sisekliimaklass muutus vähemalt ühe parameetrite kombinatsiooniga olukorras. On näha, et suur tundlikkus algoritmi parameetritele ei ole tugevalt indikatiivne sisekliimaklassi muutumisele. Analoogselt: kuulumine klastrisse, kus tundlikkus korrektsioonialgoritmi parameetritele on madal, ei taga, et sisekliimaklass korrektsiooni käigus ei muutuks. Samas, jooniselt 6.10 avaldub, et üle kogu andmestiku esineb siiski mõningane seos algoritmi parameetrite ja keskmise sisekliimaklassi väärtuse vahel. Nende nähtuste koosseisest järeldub, et sisekliimaklassi muutumise põhjused jagunevad kaheks:

- Parameetritest vähesõltuv: esineb suur baastasemete erinevus pikema ajaperioodi jooksul.
 - Sellised ruumid paiknevad põhiklastris, kuid ei asu hästi kalibreeritud andurite teljel.
- Parameetritele tundlik: kahe sisekliimaklassi piiril paiknevad tsoonid. Kõige kitsama ppm-vahemikuga on sisekliimaklass II (250 ppm). Sellega oli kooskõlas ka sisekliimaklasside üleminekumaatriksite analüüs, kus võis märgata parameetrite muutmisel kõige sagedasemalt üleminekuid üle II klassi piiride. Rohkem kui 70% kõigist üleminekutest toimusid klasside I ja II vahel.
 - Sellised ruumid ei paikne ilmingimata väljaspool põhiklastrit, kuna sisekliimaklasside piiridel paiknevad tsoonid ei vaja suurt muutust baastasemes, et põhjustada muutust klassifitseeringus. Tuleb meele pidada, et indikaatoriks on keskmine baastasemete erinevus, mitte sisekliimaklassi erinevus.



Joonis 6.10 Kõigi andurite keskmine sisekliimaklass erinevatel BLC parameetritel

6.2.3 Optimaalsete parameetrite valik

Kvantitatiivne valideerimismeetod osutus ebastabiilseks ja mitteskaleeritavaks (kirjeldatud lisas 1). Selle asemel rakendatakse kvalitatiivset optimeerimismetoodikat, tuginedes heuristilistele meetoditele nagu *Pareto* printsiip ja *Occam'i* parsimoonia. Pragmatilisest aspektist arvestatakse ka töö raames kogutud informatsiooniga:

- mõõteandmete karakteristika varieerub suures ulatuses ruumide erinevate kasutusprofiilide ja ventilatsioonilahenduste tõttu;
- mõõteandmete probleemide karakteristika ja esinemissagedus varieeruvad suures ulatuses erinevate andurimudelite lõikes;
- uuemad andurimudelid võivad juba sisaldada automaatset BLC algoritmi.

Pareto printsiip

Üldpõhimõte: põhjused, sisendid ja tööpanused jagunevad kahte kategooriasse: enamik, millel on tühine mõju ja väike vähemik, millel on valitsev mõju.

Tõlgendus töö kontekstis: tuleks keskenduda ilmselgete ja mõjusate ehk pidevate, lihtsasti piiritletavate ja suure amplituudiga vigade korrigeerimisele.

Analüüs:

Pareto printsiibil põhinev väide	Väite tõestus uuringu põhjal
Parameetrite ülitäpne optimeerimine ei ole mõistlik. Tuleb valida parameetrid, mis on minimaalselt piisavad mõjusate vigade korrigeerimiseks.	Väited tuginevad eelpoolkirjeldatud analüüsile „mõju sisekliimaklassile“ - sisekliimaklassi ennustusatavus parameetrite valiku põhjal ei ole suur.
Raskesti parandatavad või erilist lähenemist vajavad juhtumid ei peaks kuuluma üldistava korrektsioonialgoritmi töövõttu. Kaootiliselt käituvad andurid tuleb asendada funktsioneerivatega.	Töö algsemates staadiumites prooviti selliseid juhtumeid lahendada parameetrite käsitsi optimeerimise käigus, kuid selgus, et kaootilise karakteristikaga andmeridade parandamine on üldisemas plaanis algoritmi parameetrite ülesobitamine (<i>overfitting</i>) ning üldistusvõime kahaneb.

Occam'i parsimoonia

Üldpõhimõte: võrdse seletusjõuga teooriate (lahenduste) puhul tuleks eelistada vähima keerukusega teooriat (lahendust).

Tõlgendus töö kontekstis: lühema arvutusakna ja madala protsentiili kasutamine nõuavad rohkem eeldusi, millele vastavust on raske tõendada või vastavuse korral

väheneb rakendatava korrektsioonimeetodi tõsiseltvõetavus probleemi tõendus põhisel lahendamisel. Ei tasu eeldada, et valdav osa mõõteandmetest on vigased ja vajavad suurt korrektsiooni.

Analüüs:

Lühema arvutusakna rakendamise eeldused	Eeldustele vastavus uurigu andmestikus
<p>Baastase muutub ööpäevas (vm lühikese perioodi jooksul) olulisel määral ja vajab kindlasti korrektsiooni, et andmestik oleks kasutuskõlblik.</p>	<p>Anaüüsitud andmetest avaldusid ööpäevade lõikes mõningatel anduritel korrapäratud käitumismustrid, samas oli sellistes olukordades raske tegelikku baastaset hinnata ja antud korrektsioonimeetodi rakendamise alused ei oleks tugevad.</p>
<p>Ööpäeva (vm lühikese perioodi jooksul) saavutatakse alati korrektne baastase. See omakorda eeldab piisavat ventileerimist kasutusvälistel aegadel – ventilatsioonisüsteem ei tohi liiga vara välja lülituda.</p>	<p>Analüüsitud andmetest avaldus, et baastase saavutatakse enamikul juhtudest, kuid baastaseme mittesaavutamine ei ole erand, vaid moodustab märkimisväärse osa mõõteandmestikust.</p>
<p>Andurid on süstemaatiliselt üle- või alakalibreeritud (tulenevalt keskmistatud $\overline{\Delta_{bl}}$ väärtustest joonisel 6.11).</p>	<p>Joonisel 6.11 on esitatud kõigi andurite keskmine baastaseme muutus $\overline{\Delta_{bl}}$ antud parameetrite kombinatsioonidel. Mida sinisem või punasem on ruut, seda tugevam on eeldus, et andurid on süstemaatiliselt üle- või alakalibreeritud. Väidet ei saa ümber lükata, aga selle paikapidavuse tõenäosus väheneb 0-korrektsioonist (hallikad ruudud) eemaldudes, kuna pole tõestatud, et baastaseme vigade jaotus oleks tugevalt nihkes.</p>

Madalama protsentiili rakendamise eeldused	Eeldustele vastavus uurigu andmestikus
Esineb põhjendatud vajadus kasutada madalat alumise protsentiili väärtust.	Oma loomult on sellised põhjused seotud kesise väärtuste koonduvusega baastaseme ümber – suurema protsentiili kasutamisel oleks väike mõju, kui baastaseme juurde koondunud väärtuste hulk oleks piisavalt suur. Kesine koondumine tuleneb kas liiga väikesest arvutusakna suurusest (akna sees ei saavutata baastaset) või korrapäratust andmete dünaamikast. Sellistel juhtudel võib olla põhjendatud madalama protsentiili kasutamine, kuid ebastabiilne andmestik viitab ka suurele määramatusetele korrektsiooni täpsuse osas ja muudaks algoritmi töö viljakuse küsitavaks.
Mõõteandmed ei sisalda (olulisel määral) baastasemest madalama väärtusega erindeid.	Anaüüsitud andmestikus oli vastavus eeldusele suuresti olemas, kuid esines arvestataval määral erandjuhte.

Eelneva analüüsi matemaatilisel formuleerimiseks esitame kokkuvõtlikult kaks põhiprintsiipi, mille järgi saab piiritleda piisava rahuldavuse kriteeriumi lahendite huga määramiseks:

- 1) $\overline{\Delta}_{bl}$ tõenäolisem väärtus olemasolevate teadmiste põhjal on 0. Kõigi andurite keskmises maatriksis M valitakse lahendite hulk piiritletusega 0 ± 10 ppm.

$$S_1 = \{[p_{low}, t_{win,h}] \mid -10 \leq \overline{\Delta}_{bl}[p_{low}, t_{win,h}] \leq 10\} \quad (\text{Valem 6.1})$$

- 2) Alumise protsentiili p_{low} eesmärk on eemaldada minimaalne vajalik hulk tõenäolisemaist baastasemest madalamaid anomaalseid andmepunkte. Analüüsid $\overline{\Delta}_{bl}$ väärtuste tundlikkust p_{low} suhtes iga $t_{win,h}$ puhul selgub, et suuremate akende puhul esineb madalatel protsentiilidel suur tundlikkus (Joonis 6.12), samas 48 h ja lühematel akendel (Joonis 6.13) selline osa sisuliselt puudub. Tundlikkuse järsk muutus või selle puudumine vihjab erineva iseloomuga jaotustele liikuva akna ulatuses. Tundlikkuse muutumist kirjeldab matemaatilisel II diskreetne tuletis iga fikseeritud $t_{win,h}$ puhul:

$$\Delta^2 \overline{\Delta}_{bl}[p_{low,k}] = \overline{\Delta}_{bl}[p_{low,k+1}] - 2\overline{\Delta}_{bl}[p_{low,k}] + \overline{\Delta}_{bl}[p_{low,k-1}] \quad (\text{Valem 6.2})$$

Parameetrite kombinatsioonide suure muutumistundlikkusega piirkonnas domineerivad anomaalsed mustrid ja erandid, mille järgi baastaset ennustada ei tohiks või on ennustamine äärmiselt suure määramatusega. Analüüsidest kõigi olemasolevate muutumiskõverate iseloomu (kasutades kõrgema resolutsiooniga maatriksit M kui joonisel 6.11), valitakse $\Delta^2 \overline{\Delta_{bl}}$ ülemiseks lubatavaks piiriks 0,04. Siit avaldub ka teise lahendite hulga määratluskriteerium:

$$S_2 = \{(p_{low}, t_{win,h}) \mid \Delta^2 \overline{\Delta_{bl}} \geq 0,04\} \quad (\text{Valem 6.3})$$

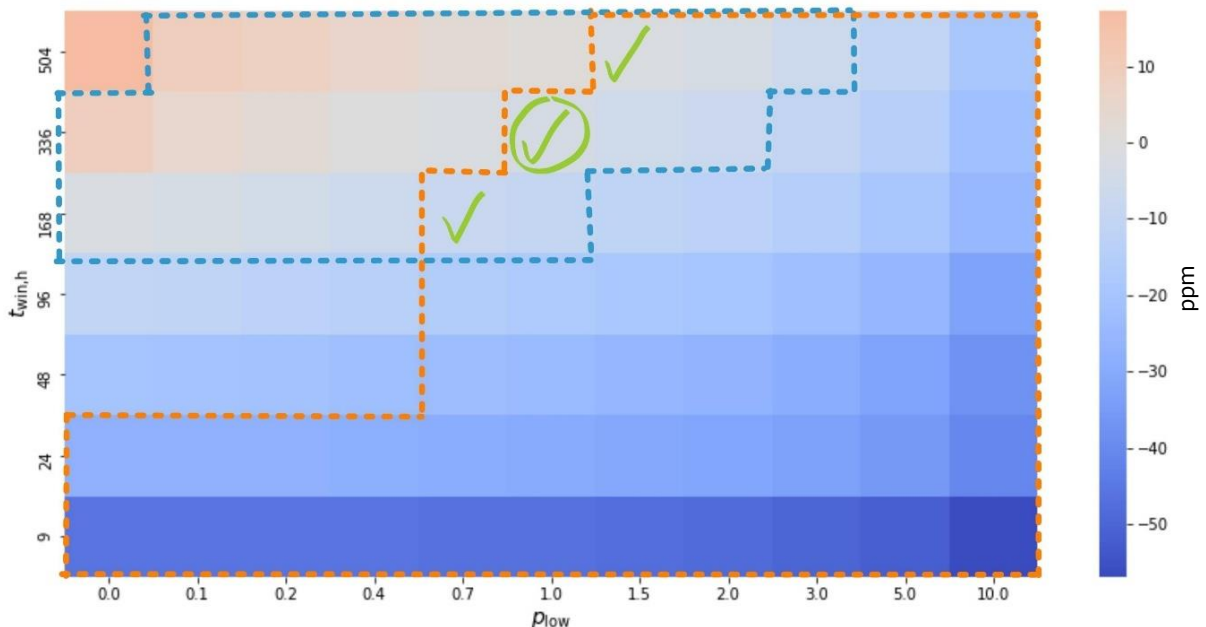
3) Leiame lahendite S_1 ja S_2 ühisosa järgi uue lahendite hulga S_3 :

$$S_3 = S_1 \cap S_2 \quad (\text{Valem 6.4})$$

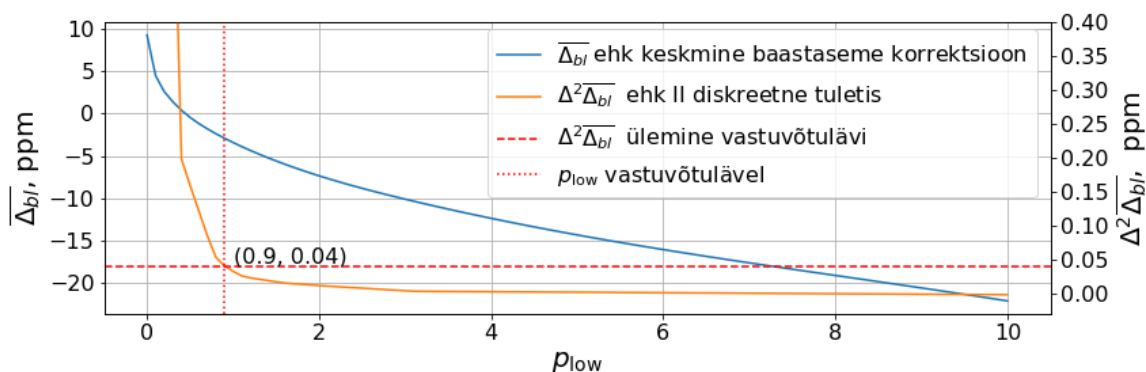
4) Lahendite hulk S_3 on endiselt liiga suur. Kitsendame valimit, teadvustades, et konstantse muutumistundlikkusega piirkonnas $\Delta^2 \overline{\Delta_{bl}} \approx 0$ on andmete jaotus ühtlustunud ning iga täiendava protsentiili suurendamisega eemaldatakse andmeid, mis suure tõenäosusega ei vaja eemaldamist. Seega tuleb valida minimeerida p_{low} väärtus iga $t_{win,h}$ puhul hulga S_3 ulatuses:

$$S = \left\{ \min_{p_{low}} \forall t_{win,h} \in S_3 \right\} \quad (\text{Valem 6.5})$$

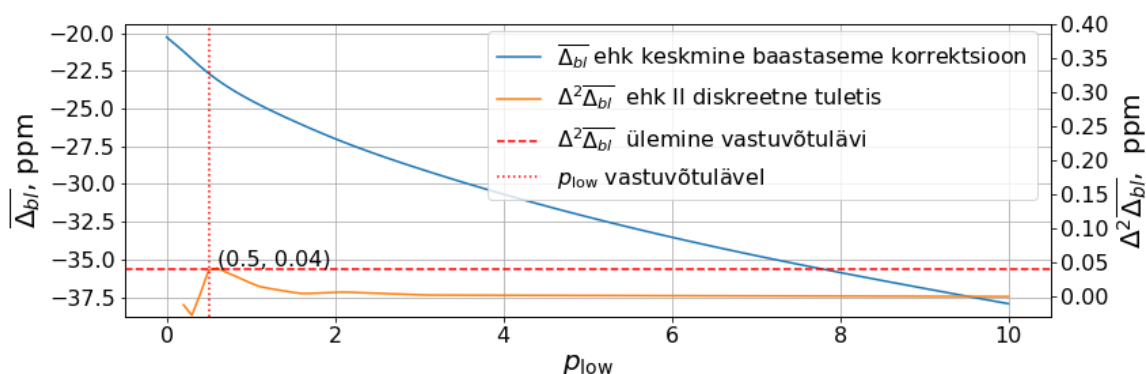
Joonisel 6.11 on kujutatud lahendite hulk S roheliste märgistega. Sinine katkendjoon piiritleb hulga S_1 ning oranž katkendjoon piiritleb hulga S_2 . Autori valitud ainulahend hulgast S on märgitud rohelse ringiga.



Joonis 6.1 Maartiks M . Kõigi andurite keskmine $\overline{\Delta_{bl}}$ on kujutatud värviskaalal. Punktiirjoontega on märgitud optimeerimisülesande kriteeriumitele vastavused. Roheliselt on märgitud optimaalsed lahendid



Joonis 6.12 p_{low} optimeerimiskriteerium $\overline{\Delta_{bl}}$ muutumisintensiivsuse järgi kui $t_{win,h} = 336$



Joonis 6.13 p_{low} optimeerimiskriteerium $\overline{\Delta_{bl}}$ muutumisintensiivsuse järgi kui $t_{win,h} = 48$

Esitatud kriteeriumite põhimõttelist korrektsust tõestab ka nende omavaheline nõustuvus lühemate akende $t_{win,h} < 168 h$ puhul:

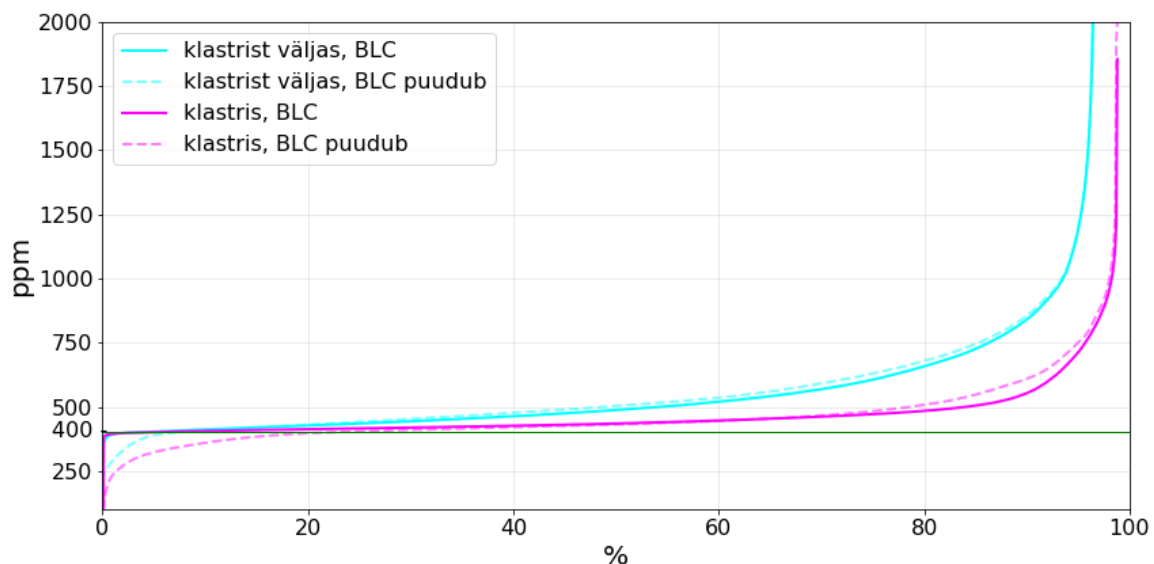
- $\overline{\Delta_{bl}}$ on ainult negatiivne ega asu S_1 piirkonnas, viidates ülekorrektsioonile
- $\Delta^2\overline{\Delta_{bl}}$ järgi on suure tundlikkusega piirkond pea olematu ning domineerib lineaarsem $\overline{\Delta_{bl}}$ muutumistrend ($\Delta^2\overline{\Delta_{bl}} \approx 0$), viidates samuti ülekorrektsioonile.

6.2.4 Optimeeritud parameetritega korrektsioonialgoritmi mõju tulemustele

Selgus, et BLC algoritmi tuleks väljundi kõrge usaldatavuse saavutamiseks rakendada ainult korrektsiooniks sobilikele ruumidele, mille tundlikkus algoritmi parameetrite suhtes on madal. Samas pole töö raames läbi viidud andurite tundlikkuspõhine eristamine praktikas lihtsasti teostatav ja minimaalse sekkumise põhimõtte parameetrite valikul on põhjendatud. Peatükis 6.2.2 esitatud aegridade graafikutel on märgitud kahe optimaalse parameetrite kombinatsiooniga teostatud BLC koos originaalse andmeregaga.

Kestuskõverad

Joonisel 6.14 on esitatud kõigi andurite kokku koondatud andmestiku kestuskõverad baastaseme korrektsioonialgoritmi ja klastrisse kuuluvuse mõju demonstreerimiseks.



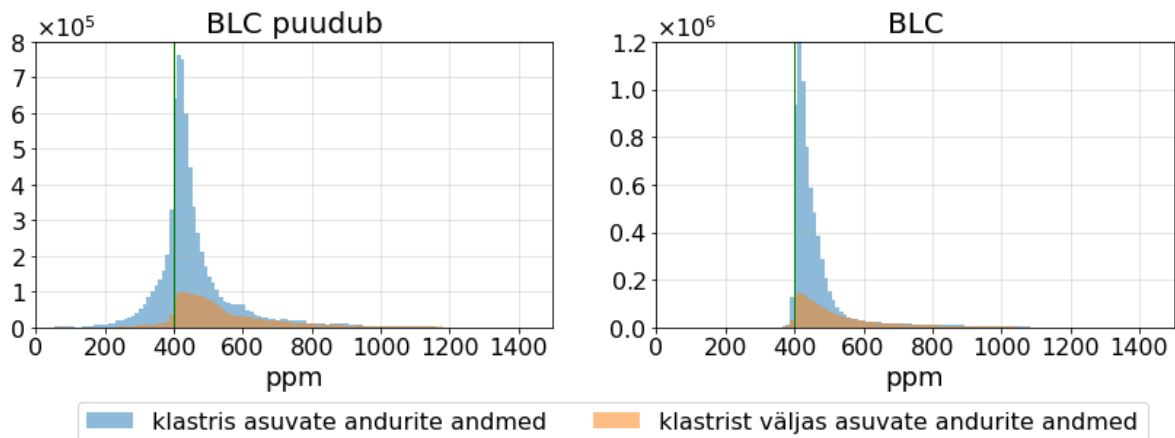
Joonis 6.14 BLC algoritmi ja klastrisse kuuluvuse mõju mõõteandmetele (ilma kasutusaegade määramiseta)

Graafikult avaldub selge erinevus klastris ja klastrist väljas asuvate väärtuste vahel. On näha, et suure BLC parameetrite tundlikkusega aegread sisaldavad rohkem kõrgemaid ppm-väärtuseid.

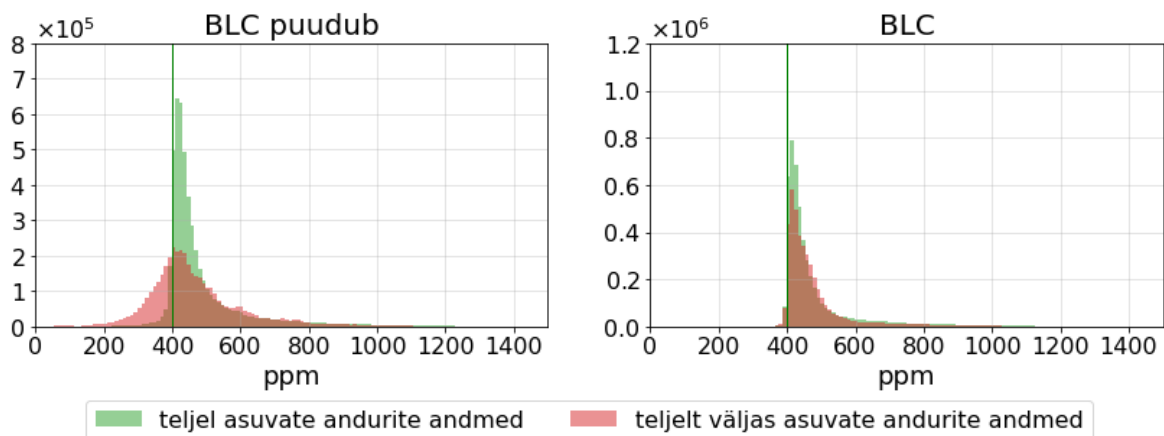
BLC algoritmi mõju avaldub tugevalt madalamatel väärtustel, mis on alla teoreetilise piiri 400 ppm. Samuti on märgata mõningast korrektsiooni klastris asuvate hoonete väärtuste vahemikus 75...95%.

Histogrammid

Järgnevad histogrammid illustreerivad BLC algoritmi mõju mõõteandmete jaotusele, eristades tulemusi klatri ja telje määratluste alusel. Diagrammidelt nähtub, et klastris ja teljel asuvate andurite andmed on jaotatud ühtlasemalt, samas kui määratluste piiridest väljaspool asuvate andurite mõõteandmete jaotused on ebakorrapärasemad. Pärast BLC algoritmi rakendamist on jaotused ühtlustunud ning muutunud ühepoolseks, nihutades kõik alla 400 ppm paiknevad väärtused realistlikku näitade vahemikku.



Joonis 6.15 BLC algoritmi ja klastrisse kuuluvuse mõju illustreerivad histogrammid

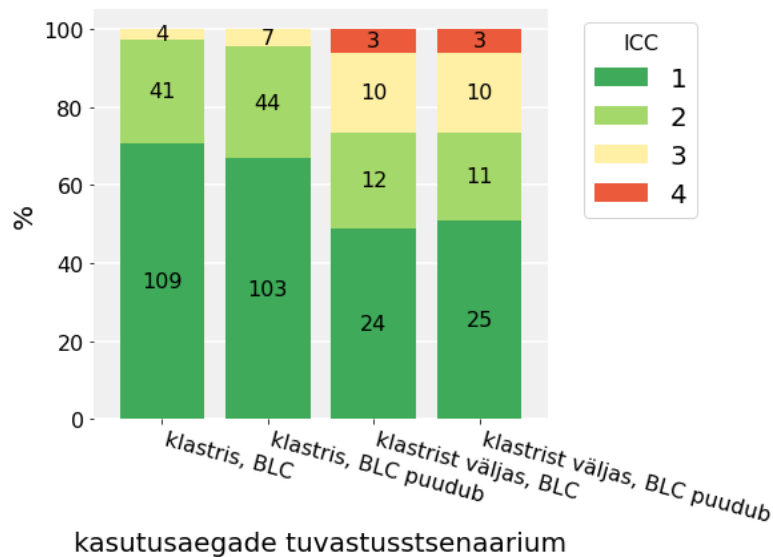


Joonis 6.16 BLC algoritmi ja teljele kuuluvuse mõju illustreerivad histogrammid

Mõju sisekliimaklassidesse jaotumisele

Joonisel 6.17 on esitatud sisekliimaklassidesse jaotuvus. Stsenaariumid on analoogsed eelnevate kestuskõverate stsenaariumitele, kuid sisekliimaklasside arvutamiseks on määratud ka kasutusajad.

Sarnaselt kestuskõveratele on näha, et klastrist väljas olevad andurid saavutavad kehvemad sisekliimaklassid. See annab täiendava motivatsiooni uurida klastrist väljas asuvate ruumide kasutusprofiilide, ventilatsioonisüsteemide ja andurite töö adekvaatsuse tagamist, et selgitada välja kesisemaid tulemusi põhjustavad tegurid ja mõõtetulemuste ning andmetötlusalgoritmide tõsiseltvõetavus.



Joonis 6.17 Sisekliimaklassidesse kuuluvate ruumide arv 100% skaalal, demonstreerimaks BLC algoritmi ja klastrisse kuuluvuse mõju

6.3 Kasutusaegade tuvastamine

6.3.1 Algoritmi töö ülevaade ja piirangud

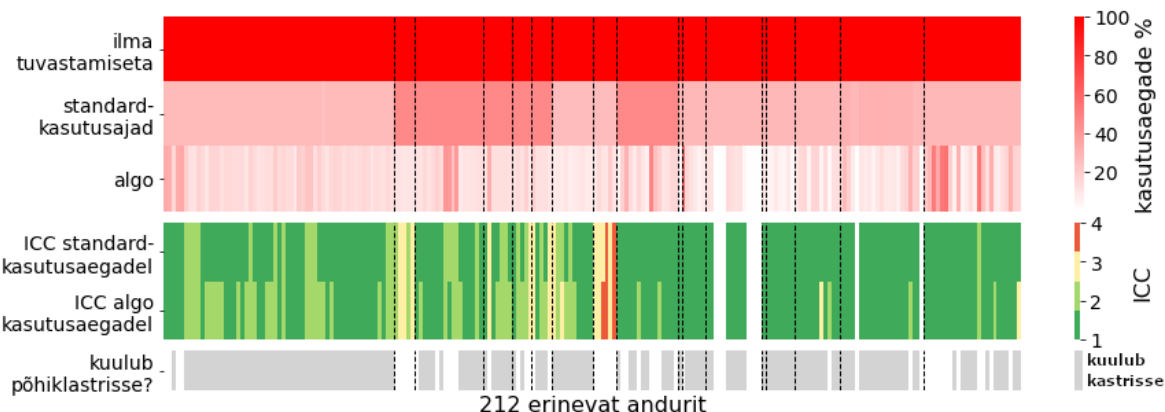
Aegridade graafikutel peatükis 6.2.2 on esitatud mõned näidised algoritmi poolt tuvastatud kasutusaegadest. Tuvastatud kasutusajad on märgitud helepunaste perioodidena.

Algoritmi tööd analüüsid avaldus põhiklastrist väljas olevate andurite puhul suur risk valepositiivseteks tuvastusteks. Valepositiivsed tuvastused tulenesid ebastandardsetest mustritest aegridades (Joonis 6.7 ja Joonis 6.8), millega algoritmi arendamise käigus ei olnud arvestatud. Selgus, et kasutusaegade algoritmiline tuvastamine on kompleksne ülesanne, mille edukaks läbiviimiseks suure varieeruvusega andmestiku ulatuses tuleb suurendada algoritmi keerukust erandjuhtude käsitlemiseks. Üks lahendustest oleks kasutada komplekssemaid masinõppemeetodeid, mis suudavad eristada anomaalsed käitumismustrid tüüpilistest.

Antud töös on kasutusaegade tuvastuse analüüsi eesmärk demonstreerida olemasoleva standardmetoodika järgse sisekliimaklasside arvutuse tundlikkust kasutusaegade määramisele. Seega kasutatakse edaspidises analüüsis vaid defineeritud põhiklastrit, kus on väga suur tõenäosus algoritmi ootuspäraseks toimimiseks. Klaster sisaldab ligikaudu 66,7% kõigist anduritest.

Kasutusaegade tuvastuse ja sisekliimaklasside arvutuste lõpptulemused andurite ja hoonete lõikes koondab kokku joonis 6.18. Andurite järjekord horisontaalteljel, sh hoonete eraldised, kattuvad eelnevate sarnaste joonistega, kus horisontaalteljele on

märgitud „212 erinevat andurit“. Järgnevatel peatükkides võrreldakse tuvastusstsenariumeid erinevatest aspektidest.



Joonis 6.18 Kasutusaegade tuvastuse ja sisekliimaklasside tulemuste ülevaatlik maatriks

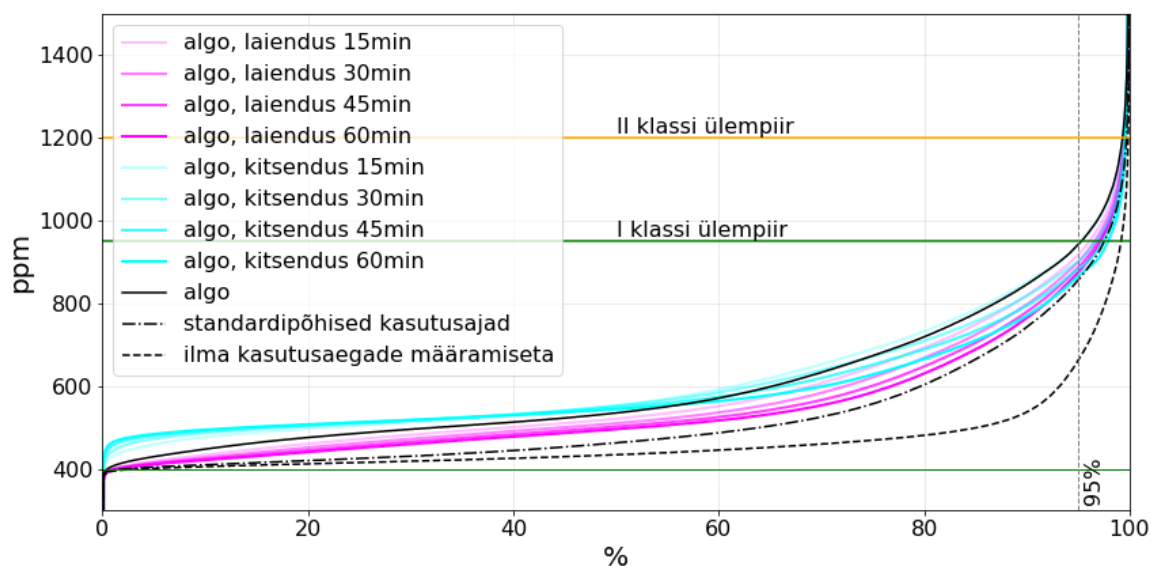
6.3.2 Tuvastusstsenariumite võrdlus

Kestuskõverad

Joonisel 6.19 on esitatud põhiklastri andmestiku kestuskõverad erinevate kasutusaegade stsenaariumite puhul:

- algoritmiliselt tuvastatud kasutusajad, sh laiendused ja kitsendused
- standardkasutusel põhinevad, staatilised kasutusajad
- ilma kasutusaegade määramiseta

Iga stsenaariumi keskmise sisekliimaklassi saab jooniselt määrata, kui leida 5% tolereeritud ületuspiiri (vertikaalne 95% joon) ja kestuskõvera lõikepunkti asukoht sisekliimaklassipiiride suhtes. Näiteks, kui lõikepunkt on I klassi piirist kõrgemal, saavutatakse II klass.



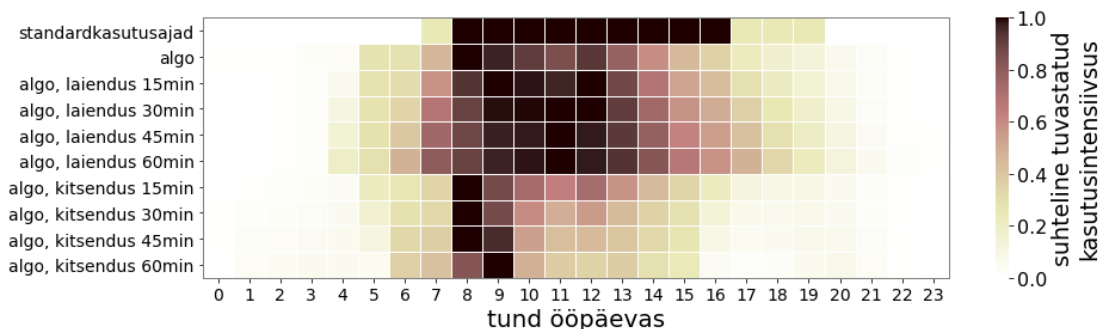
Joonis 6.19 Klastri asuvate andurite koondandmete kestuskõverad erinevates kasutusaegade tuvastamise stsenaariumites

Jooniselt nähtub, et keskmine sisekliimaklass sõltub kasutusaegade määramise täpsusest. Algoritmiliselt tuvastatud kasutusaegasid laiendades muutub keskmine sisekliimaklass klassist II klassi I. Ekstreemjuhud kasutusaegade laiendamisest on standardkasutuse ja ilma kasutusaegade tuvastamiseta stsenaariumid. On näha, et need stsenaariumid kaasavad palju madala väärtusega kasutusajaväliseid mõõteandmeid, mis muudavad tulemused liialt optimistlikuks.

Ajaline kattuvus

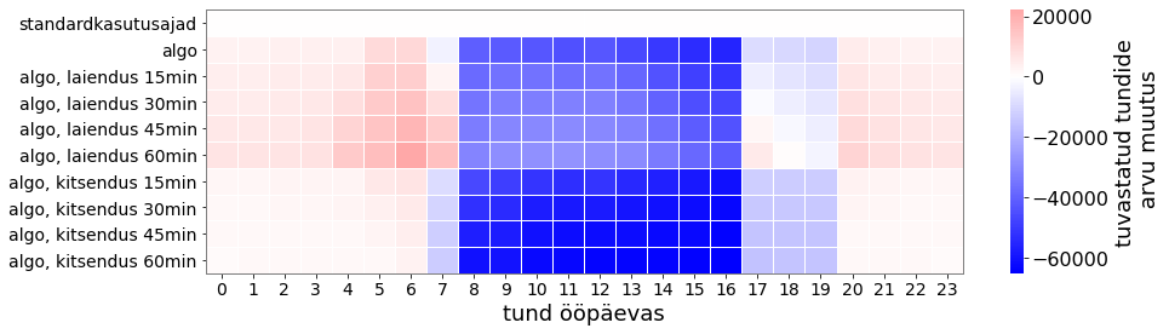
Joonisel 6.20 on esitatud erinevates stsenaariumites tuvastatud tundide suhteline intensiivsus ööpäeva lõikes. „Suhteline“ tähendab, et iga stsenaariumi tuvastatud kasutustunnid on normaliseeritud 0...1 vahemikku lähtuvalt minimaalsest ja maksimaalsest väärtusest.

Algoritmiliselt tuvastatud kasutusajad on ööpäeva lõikes hajusamad, samas keskmine suhteline tuvastatud kasutusintensiivsus alaneb juba enne standardkasutuse tundide lõppu. Algoritmi tuvastusperioodide laiendamisel päevase profiili ajaline määratlus hägustub ning kaasatakse ka rohkem standardse kasutusaaja väliseid tunde. Tuvastusperioodide kitsendamisel langeb suurem rõhk tööpäeva algusele, kus tõenäoliselt esinevad pikemad tuvastatud kasutusetaapid.



Joonis 6.20 Erinevate kasutusaegade stsenaariumite suhteline tuvastatud kasutusintensiivsus ööpäeva lõikes

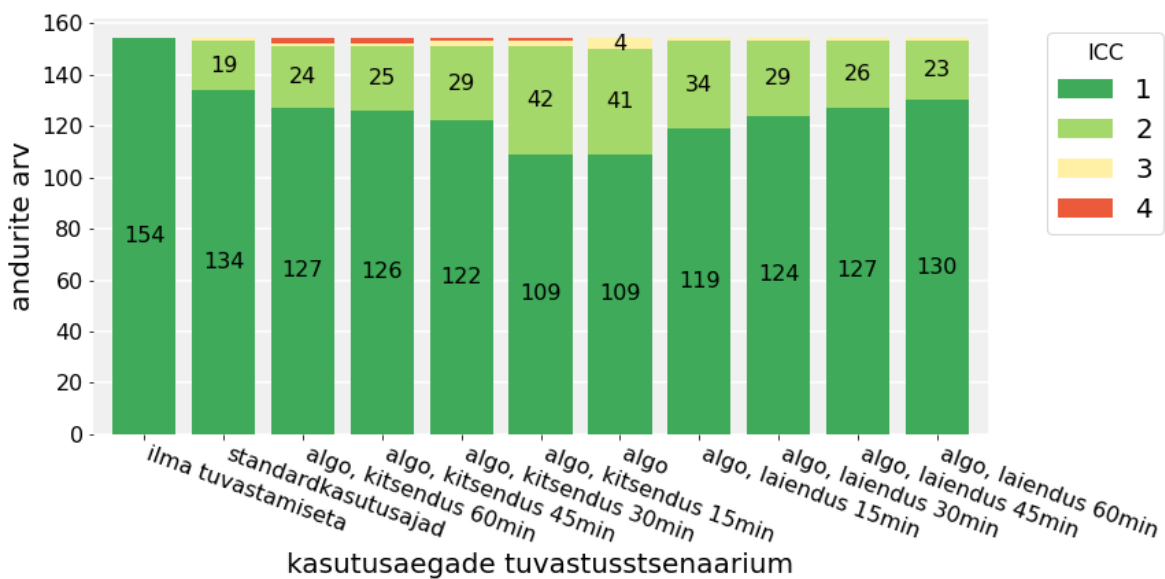
Joonisel 6.21 on esitatud tuvastatud tundide arvu erinevus võrreldes standardkasutusaegade stsenaariumiga. Kooskõlas joonisega 6.18 avaldub, et algoritmiliselt tuvastatud kasutusajad on märkimisväärselt lühemad, moodustades keskmiselt vaid ligikaudu 40% standardijärgsetest kasutusaegadest.



Joonis 6.21 Erinevate kasutusaegade stsenaariumite tuvastatud tundide arvu muutus võrreldes standardkasutusaegadega

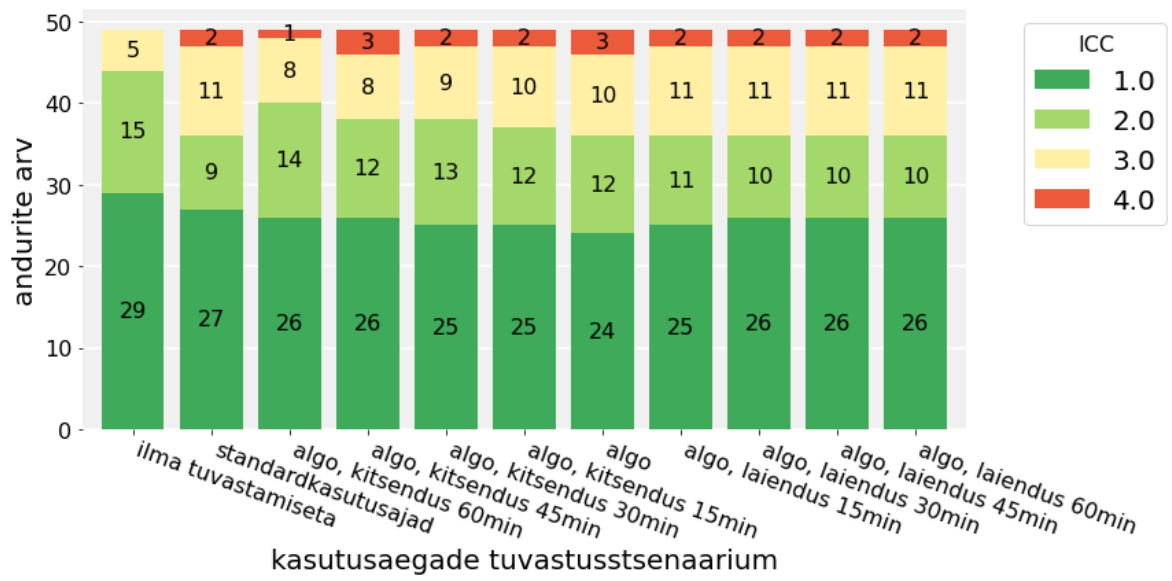
Mõju sisekliimaklassidesse jaotumisele

Kõigi klastris asuvate andurite jaotuvus sisekliimaklassidesse on kuvatud joonisel 6.22. Võib märgata, et kõige kriitilisemad tulemused avalduvad kasutusaegade tuvastuse algoritmi algparameetrite puhul (stsenaariumis „algo“) ning 15-minutilise kitsenduse puhul. Kuna need stsenaariumid sarnanevad ka ööpäeva lõikes joonisel 6.20, võib järeldada, et mõõdukalt alahinnatud kasutusajad võivad siiski sobida sisekliimaklasside arvutamiseks.



Joonis 6.22 Klastris asuvate ruumide jaotuvus sisekliimaklassidesse erinevates kasutusaegade tuvastamise stsenaariumites

Joonisel 6.23 on näidatud klastrist väljas asuvate andurite jaotuvus sisekliimaklassidesse. Puudub tugev seos kasutusaegade laiendamise ja arvutusliku sisekliimaklassi vahel, viidates muudele teguritele, sh anomaaliatele andmestikus, mis mõjutavad tulemusi eraldiseisvalt.



Joonis 6.23 Klalstrist väljas asuvate ruumide jaotuvus sisekliimaklassidesse erinevates kasutusaegade tuvastamise stsenaariumites

7. JÄRELDUSED

CO₂ andurite kasutus hoonetes on tõusutrendis ja tõenäoliselt hoogustub veelgi, tulenevalt kõrgenevatest nõuetest hoonete sisekliimale Euroopa Liidus. Kuigi on olemas standardmetoodikad, mis annavad juhised CO₂-põhiseks õhukvaliteedi hindamiseks, tuleb praktikas arvestada erinevate probleemidega, mis pärsivad soovitud, adekvaatse väljundi saavutamist.

Järgnevalt esitatakse töö käigus analüüsitud teaduskirjanduse ning 333 aasta kogupikkusega mõõteandmete põhjal kokku pandud ülevaatlik pingerida peamistest probleemidest vastavalt nende mõjukusele. Mõju ning esinemissagedus on autori empiirilised hinnangud, üldistades töö käigus ammendatud informatsiooni.

Tabel 7.1 Ülevaatlik pingerida peamistest probleemidest ning nende mõjust tulemustele

Probleemi kirjeldus	Mõju tulemustele	Esinemissagedus
Andmeedastussüsteemi lühikesed, juhuslikud, lokaalsed häiringud	Väga tugev	Kõrge
Andmeedastussüsteemi pikad, sünkroonsed, mastaapsed häiringud	Väga tugev	Kõrge
Eksimused andurite märgistamisel ja andmebaasi kirjutamisel	Väga tugev	Arvestatav
Kasutusaegade ülehindamine rakendades standardkasutusaegasid	Tugev	Kõrge
Anduri vigane kalibreering	Tugev	Kõrge
Anduri düsfunktsionaalne integreeritud tarkvara või füüsiline komponent	Tugev	Arvestatav
Mõõteandmete ebastandardne dünaamika	Tugev	Arvestatav
Anduri deklareeritud täpsusele mittevastavus ja sõltuvused välisteguritest	Keskmine	Arvestatav
Anduri ebakvaliteetne paigutus	Arvestatav	Teadmata
Kasutusaegade alahindamine rakendades lihtsat kasutusaegade tuvastusalgoritmi	Madal	Arvestatav
Anduri kalibreeringu nihkumine ajas	Madal	Madal

Töösse kaasatud andurite ja andmeedastussüsteemide valim ei ole piisavalt laialdane, et tuvastatud probleeme ekstrapoleerida kõigile turul leitavatele toodetele. Antud valimi probleeme tuleb teadvustada, et suhtuda monitoorimisahela komponentide ja tehniliste lahenduste valikusse kriitiliselt, vältimaks sarnaseid probleeme uutes süsteemides.

Andmeedastussüsteemid ja andmete pidevus

On võimalik eristada erineva töökindlusega andmeedastussüsteemide lahendusi:

- süsteemid, kus andmeedastus on pidev ja tõrkeid ei esine;
- süsteemid, kus esinevad lühikesed, juhuslikud, lokaalsed häiringud;
- süsteemid, kus esinevad pikad, sünkroonsed, mastaapsed häiringud.

Andmeedastussüsteemide töökindlus on kriitilise tähtsusega hoonete kaugjälgimisel, sealhulgas CO₂-põhise siseõhu kvaliteedi hindamisel. Tuleb pöörata tähelepanu andmeedastusahela komponentide ja suhtlusprotokollide ühilduvusele ning võrgu stabiilsusele ja läbilaskevõimele.

Andurite töökindlus ja andmete karakteristik

Andurite töökindlus võib erineda märkimisväärselt. Tundlikkus baastaseme korrektsioonialgoritmi parameetrite muutmise suhtes aitab rühmitada andureid mõõteandmete karakteristik alusel:

- andurid, mille kalibreering on ning püsib ajas suurepärase. Andmete dünaamika on ootuspärane ja hästi tõlgendatav. Tõenäoliselt on sellistes andurites hästi funktsioneerivad olemasolev(ad) korrektsioonialgoritm(id).
- andurid, mille kalibreering on kehv ning võib muutuda ajas. Andmete ajaline dünaamika on siiski hästi tõlgendatav. On kõrge potentsiaal lihtsa korrektsioonialgoritmi edukaks rakendamiseks ning kasutusaegade tuvastamiseks.
- andurid, mille kalibreering tundub olevat korrektne, kuid dünaamika on raskestitõlgendatav. Tuleb välja selgitada, kas ebastandardised mustrid CO₂ tasemes kajastavad tegeliku füüsilise keskkonna eripärasid või on tegu düsfunktsionaalsete anduritega. BLC algoritmi rakendamine ei anna lisandväärtust ning kasutusaegade tuvastamine on nüansirohkem.
- andurid, mille kalibreeringut ja mõõtetulemuste tõesust on raske hinnata tulenevalt anomaalsest ja raskesti tõlgendatavast dünaamikast. On madal potentsiaal korrektsioonialgoritmide edukaks rakendamiseks ning kasutusaegade tuvastamiseks. Tuleb täiendavalt uurida anomaalsete mustrite esinemissagedust alternatiivse valimiga andmestikus ning leida tekkepõhjused.

Baastaseme korrektsioon

Vaatamata arengule NDIR andurite tehnoloogias ja integreeritud tarkvaras, pole nihkuva baastaseme probleemid laiemas mastaabis veel täielikult lahendatud. Et suurendada mitmekesise andurimudelite valimiga mõõteandmete usaldusväärsust, tuleks rakendada baastaseme korrektsioonialgoritmi. Algoritm peaks keskenduma ilmselgete ja mõjusate vigade parandamisele, vältides ülekorrektsiooni. Selle

saavutamiseks tuleks korrigeerimiseks kasutada suhteliselt laia liikuvat arvutusakent (näiteks pikkusega 2 nädalat). Baasase tuleks määrata minimaalse võimaliku alumise protsendi väärtuse järgi (näiteks 1%), et jätta välja anomaaliatest või mõõtemääramatusest tingitud andmepunktid, mis asuvad madalamal kui baasnivoos.

Kasutusaegade tuvastamine

Tegelikud kasutusajad ei kata kunagi standardkasutusaegade määratletud perioode täielikult. Algoritmiliselt tuvastatud kasutusajad moodustasid töö andmestiku keskmisena vaid ligi 40% standardijärgsetest kasutusaegadest. Kasutusaegade tuvastuse algoritmi rakendamise eesmärgid saab jagada kaheks: standardijärgse sisekliimaklassi arvutamine või energia kokkuhoiust ajendatud tegelike kasutusaegade hindamine, sh tehnosüsteemide töö optimeerimine.

Sisekliimaklasside arvutamisel on oluline saavutada mõõteandmete valimi esinduslik jaotus kriitilisematel ajaperioodidel, mitte maksimaalne andmete hulk. Sellest tulenevalt sobivad ka mõõdukalt alahinnatud, kuid kõrge usaldusnivooga kasutusajad sisekliima hindamiseks. Standardijärgsed või mõnel muul moel ülehinnatud kasutusajad muudavad tulemused liialt optimistlikuks, kaasates palju kasutusajaväliseid madalaid kontsentratsioone. Sisekliima hindamise seisukohast on kasutusaegade tuvastamiseks õigustatud kasutada lihtsat ja robustset tuvastusalgoritmi, mis tähtsustab valedpositiivsete tuvastuste vältimist valenegatiivsete tuvastuste arvelt. Kõrge usaldusnivooga tuvastused aitavad rõhutada kriitilisemaid olukordi, mil liigkõrge CO₂ taseme poolt indikeeritud ohud ruumisviibijate tervisele ja heaolule on ilmsed.

Kasutusaegade ajalise tuvastustäpsuse maksimeerimine nõuaks valideeritud mõõteandmetega treenitud masinõppepõhist algoritmi, mis suudaks käsitleda ka ebastandardset CO₂ dünaamikat.

KOKKUVÕTE

CO₂ andurite kasutamine hoonetes on tõusutrendis ja tõenäoliselt hoogustub veelgi, tulenevalt kõrgenevatest nõuetest hoonete sisekliimale Euroopa Liidus. Kuigi on olemas standardmetoodikad, mis annavad juhised CO₂-põhiseks õhukvaliteedi hindamiseks, tuleb praktikas arvestada erinevate probleemidega.

Töös analüüsiti mõõteandmete massiivi, mis pärineb 17 erinevast haridushoonest 212 erinevalt andurilt, sisaldades kokku 333 aasta jagu mõõtmisandmeid. Suurel skaalal teostatud analüüsi käigus avaldusid erinevat liiki andmeedastuse häired, andurite kalibreerimisvead ja mõõteandmete ebastandardne ajaline dünaamika.

Vaatamata arengule NDIR andurite tehnoloogias ja tarkvaras, ei olnud korrektne CO₂ baastase siiski sageli tagatud ning ilmnis vajadus rakendada täiendavalt baastaseme korrektsioonialgoritmi. Algoritmi tööd optimeeriti nii, et säiliks lihtsus, usaldatavus, skaleeritavus ja oleks välistatud ülekorrektsioon.

Standardijärgsete sisekliimaklasside arvutamiseks hindamiseks on vaja määratleda ruumi kasutusajad. Ilmnes, et CO₂ signaali järgi algoritmiliselt tuvastatud kasutusajad moodustasid keskmiselt vaid 40% standardkasutusaegadest. Seega on standardkasutusaegade järgi arvutatud õhukvaliteediklass sageli liialt optimistlik: algoritmilistel kasutusaegadel kandus ca 17% anduritest madalamasse klassi. Samas avaldus, et õhukvaliteediklassi määramiseks piisab ka veidi alahinnatud, kuid kõrge usaldusnivooga kasutusaegade tuvastamisest, mis loob eeldused lihtsa kasutusaegade tuvastusalgoritmi rakendamiseks.

Usaldusväärsete CO₂ mõõtetulemuste saavutamiseks on äärmiselt oluline valida töökindel andurimudel ning paigutada see ruumis hästi segunenud õhuga piirkonda, eemale õhujugadest ja saasteallikatest. Täendava kindlustunde lisab baastaseme korrektsioonialgoritmi kasutamine, luues eeldused ka lihtsa kasutusaegade tuvastusalgoritmi ootuspäraseks tööks ning õhukvaliteedi edukaks hindamiseks.

SUMMARY

The use of CO₂ sensors in buildings is on the rise and is expected to increase further due to the heightened requirements for indoor climate in the European Union. Although there are standard methodologies that provide guidelines for CO₂-based air quality assessment, various issues need to be considered in practice.

In this study, a dataset was analyzed, comprising measurements from 17 different educational buildings with 212 different sensors, amounting to a total of 333 years of measurement data. Large-scale analysis revealed various types of data transmission errors, sensor calibration errors, and non-standard temporal dynamics of the measurement data.

Despite advancements in NDIR sensor technology and software, the correct CO₂ baseline was often not maintained, indicating the need to apply an additional baseline correction algorithm. The algorithm was optimized to ensure simplicity, reliability, scalability, and to avoid overcorrection.

For the calculation of indoor climate classes according to EN standard, it is necessary to define the room usage times. It was found that the usage times which were identified algorithmically from the CO₂ signal averaged only 40% of the standard usage times. Therefore, the air quality class calculated according to standard usage times is often overly optimistic: about 17% of the sensors fell into a lower class based on algorithmic usage times. Nevertheless, it was revealed that slightly underestimated but highly reliable usage times are sufficient to determine the air quality class, creating the prerequisites for the application of a simple usage time detection algorithm.

To achieve reliable CO₂ measurement results, it is crucial to select a reliable sensor model and place it in a well-mixed air area of the room, away from air jets and pollution sources. Additional confidence is provided by using a baseline correction algorithm, which also establishes the prerequisites for the expected operation of a simple occupancy detection algorithm and a successful assessment of indoor air quality.

KASUTATUD KIRJANDUSE LOETELU

- Alfano, F., Bellia, L., Boerstra, A., Dijken, F., Ianniello, E., Lopardo, G., . . . da Silva, M. (2010). Indoor Environment and Energy Efficiency in Schools, Part 1 Principles. rmt: *REHVA Guidebook* (lk 7-8).
- Aljas, H. K., Thalfeldt, M., & Parts, T. M. (2023). Building-wise algorithm for indoor air quality classification based on measured CO₂ data. *Proceedings of Healthy Buildings 2023 Europe*.
- Berquist, J., & Banister, C. (2020). Feasibility of Low-Cost CO₂ Sensors for Demand-Controlled Ventilation -Laboratory Chamber Testing. *ASHRAE Transactions*, 126(2).
- Boerstra, A., Raue, A., & Cheng, L. (2019). Smart monitoring of building performance with IEQ sensor network. *REHVA Journal*, 56(2), 6–12.
- Bulińska, A., Popiołek, Z., & Buliński, Z. (2014). Experimentally validated CFD analysis on sampling region determination of average indoor carbon dioxide concentration in occupied space. *Building and Environment*, 72, 319–331. <https://doi.org/https://doi.org/10.1016/j.buildenv.2013.11.001>
- CO₂meter. (4. 03 2023. a.). *What is a CO₂ Gas Detector and How Does it Work?* Allikas: <https://www.co2meter.com/blogs/news/co2-carbon-dioxide-detector>
- Deng, H.-Y., Feng, Z., & Cao, S.-J. (2018). Influence of air change rates on indoor CO₂ stratification in terms of Richardson number and vorticity. *Building and Environment*, 129, 74–84. <https://doi.org/https://doi.org/10.1016/j.buildenv.2017.12.009>
- Dinh, T.-V., Choi, I.-Y., Son, Y.-S., & Kim, J.-C. (2016). A review on non-dispersive infrared gas sensors: Improvement of sensor detection limit and interference correction. *Sensors and Actuators B: Chemical*, 231, 529–538. <https://doi.org/https://doi.org/10.1016/j.snb.2016.03.040>
- Euroopa Standardimiskomitee (CEN). (2019). *EN 16798-1:2019 - Energy performance of buildings - Ventilation for buildings - Part 1: Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting, and ac*. Brüssel: CEN.
- Fisk, W., Faulkner, D., & Sullivan, D. (2008). *A pilot study of the accuracy of CO₂ sensors in commercial buildings*.
- Fisk, W. J., Sullivan, D. P., Faulkner, D., & Eliseeva, E. (2010). *CO₂ Monitoring for Demand Controlled Ventilation in Commercial Buildings*. <https://doi.org/10.2172/983161>
- Keimel, A. (2019). *Comparison of low-cost CO₂ non-dispersive infrared (NDIR) sensors for ambient greenhouse gas monitoring* [UVM Honors College Senior Theses]. University of Vermont.
- Klepeis, N., Nelson, W., Ott, W., Robinson, J., Tsang, A., Switzer, P., Behar, J., Hern, S., & Engelmann, W. (2001). The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *Journal of Exposure Analysis and Environmental Epidemiology*, 11, 231–252. <https://doi.org/10.1038/sj.jea.7500165>
- Kurnitski, J. (2023). IEQ in the EPBD – How to set national requirements in line with revised EPBD? *REHVA Journal*, 60(6), 17–21.
- Mahyuddin, N., & Awbi, H. (2010). The spatial distribution of carbon dioxide in an environmental test chamber. *Building and Environment*, 45(9), 1993–2001. <https://doi.org/https://doi.org/10.1016/j.buildenv.2010.02.001>
- Mou, J., Cui, S., & Khoo, D. W. Y. (2022). Computational fluid dynamics modelling of airflow and carbon dioxide distribution inside a seminar room for sensor placement. *Measurement: Sensors*, 23, 100402. <https://doi.org/https://doi.org/10.1016/j.measen.2022.100402>
- Mylonas, A., Kazanci, O. B., Andersen, R. K., & Olesen, B. W. (2019). Capabilities and limitations of wireless CO₂, temperature and relative humidity sensors. *Building*

- and Environment*, 154, 362–374.
<https://doi.org/https://doi.org/10.1016/j.buildenv.2019.03.012>
- Park, J., Cho, H., & Yi, S. (2010). NDIR CO₂ gas sensor with improved temperature compensation. *Procedia Engineering*, 5, 303–306.
<https://doi.org/https://doi.org/10.1016/j.proeng.2010.09.108>
- Pedersen, T. H., Nielsen, K. U., & Petersen, S. (2017). Method for room occupancy detection based on trajectory of indoor climate sensor data. *Building and Environment*, 115, 147–156.
<https://doi.org/https://doi.org/10.1016/j.buildenv.2017.01.023>
- Pei, G., Rim, D., Schiavon, S., & Vannucci, M. (2019). Effect of sensor position on the performance of CO₂-based demand controlled ventilation. *Energy and Buildings*, 202, 109358. <https://doi.org/https://doi.org/10.1016/j.enbuild.2019.109358>
- Peng, Z., Rojas, A. L. P., Kropff, E., Bahnfleth, W., Buonanno, G., Dancer, S. J., Kurnitski, J., Li, Y., Loomans, M. G. L. C., Marr, L. C., Morawska, L., Nazaroff, W., Noakes, C., Querol, X., Sekhar, C., Tellier, R., Greenhalgh, T., Bourouiba, L., Boerstra, A., ... Jimenez, J. L. (2022). Practical Indicators for Risk of Airborne Transmission in Shared Indoor Environments and Their Application to COVID-19 Outbreaks. *Environmental Science & Technology*, 56(2), 1125–1137.
<https://doi.org/10.1021/acs.est.1c06531>
- Shrestha, S., & Maxwell, G. (2010a). *An Experimental Evaluation of HVAC-Grade Carbon-Dioxide Sensors: Part 2, Performance Test Results*. 116.
- Shrestha, S., & Maxwell, G. (2010b). *An Experimental Evaluation of HVAC-Grade Carbon-Dioxide Sensors: Part 3, Humidity, Temperature, and Pressure Sensitivity Test Results*. 116.
- Shrestha, S., & Maxwell, G. (2010c). An Experimental Evaluation of HVAC-Grade Carbon-Dioxide Sensors: Part 4, Effects of Ageing on Sensor Performance (AB-10-006). *ASHRAE Transactions*, 116.
- Shrestha, S., & Maxwell, G. M. (2009). An experimental evaluation of HVAC-grade carbon dioxide sensors -Part I: Test and evaluation procedure. *ASHRAE Transactions*, 115, 471–483.
- Stafford, T. M. (2015). Indoor air quality and academic performance. *Journal of Environmental Economics and Management*, 70, 34–50.
<https://doi.org/https://doi.org/10.1016/j.jeem.2014.11.002>
- Teng, T.-P., & Chen, W.-J. (2023). A compensation model for an NDIR-based CO₂ sensor and its energy implication on demand control ventilation in a hot and humid climate. *Energy and Buildings*, 281, 112738.
<https://doi.org/https://doi.org/10.1016/j.enbuild.2022.112738>
- Vihman, L., Parts, T. M., Aljas, H. K., Thalfeldt, M., & Raik, J. (2023). Algorithms for online CO₂ baseline correction in intermittently occupied rooms. *Proceedings of Healthy Buildings 2023 Europe*.
- Wargocki, P., Wei, W., Bendžalová, J., Espigares-Correa, C., Gerard, C., Greslou, O., ... Mandin, C. (2021). Tail, a new scheme for rating indoor environmental quality in offices and hotels undergoing Deep Energy Renovation (EU aldren project). *Energy and Buildings*, 244, 111029.
<https://doi.org/10.1016/j.enbuild.2021.111029>
- Yi, S. (2017). Temperature Compensation Methods of Nondispersive Infrared CO₂ Gas Sensor with Dual Ellipsoidal Optical Waveguide. *Sensors & Materials*, 29(3).
- Zhou, L., He, Y., Zhang, Q., & Zhang, L. (2021). Carbon Dioxide Sensor Module Based on NDIR Technology. *Micromachines*, 12(7).
<https://doi.org/10.3390/mi12070845>

LISA 1 - BAASTASEME KORREKTSIOONI VALIDEERIMINE PIIRKONTSENTRATSIOONIDE MEETODIL

Puuduvate tõeste mõõteandmete tõttu otsustati töö esialgses käsitluses rakendada baastase korrektsioonialgoritmi töö pisteliseks valideerimiseks CO₂ hajumise piirkontsentratsiooni meetodit.

Meetodi eeldusteks on täielikult segunenud õhuvahetus, konstantne õhuvooluhulk ning kasutusaegade välised, häiringuteta eksponentsiaalsed langusperioodid, mis vastavad seosele

$$CO_2[t] = a \cdot e^{-n(t-t_0)} + b \cdot (1 - e^{-n(t-t_0)})$$

kus $CO_2[t]$ on CO₂ kontsentratsioon jada elemendi indeksi t kohal;

t_0 on jada esimese elemendi indeksi number

a on CO₂ kontsentratsioon alghetkel t_0

b on tasakaalukontsentratsioon, ppm

n on õhuvahetuse kordarv, ühik vastavalt jada ajaresolutsioonile

Algoritm tuvastab sellised pikad häiringuteta langusperioodid. Seejärel on algoritmi töö järgmine:

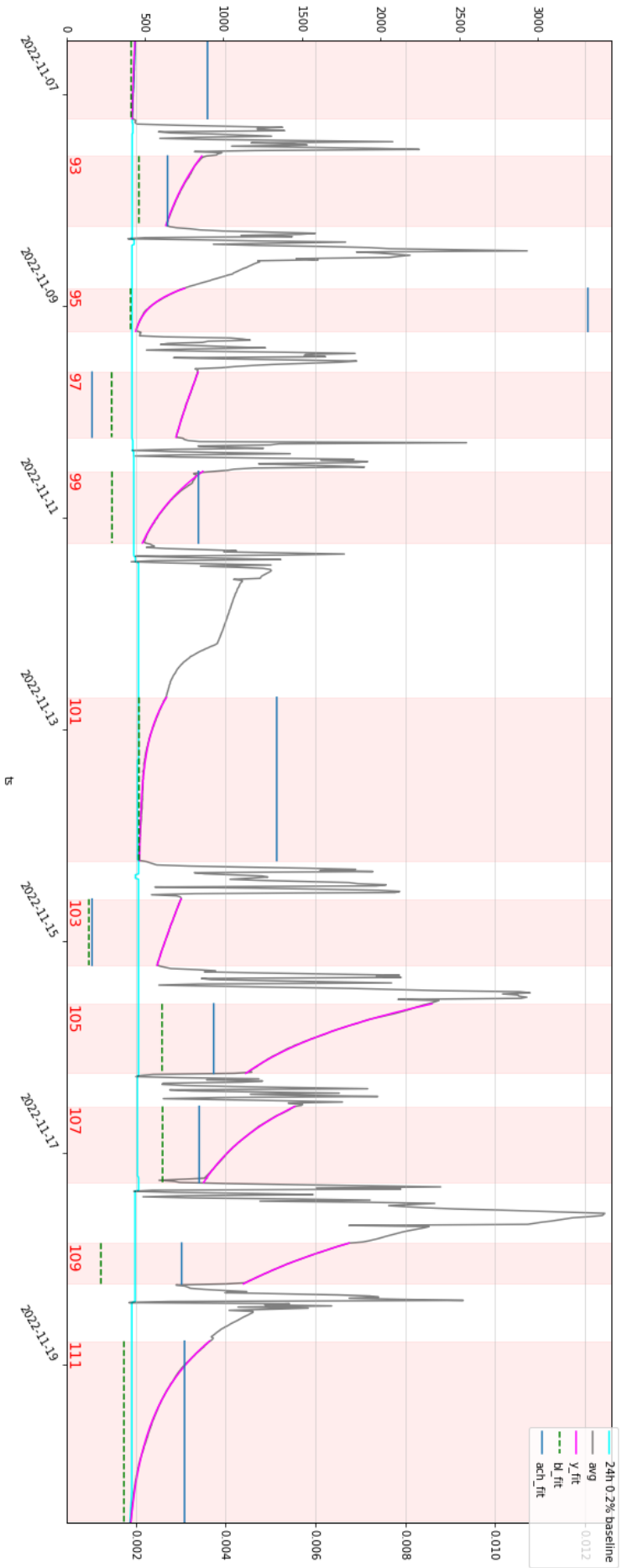
- 1) igale tuvastatud langusperioodile sobitatakse regressioonikordajad kasutades iteratiivset sobitamist (igal iteratsioonil jäetakse välja kõige halvemini sobituvad andmepunktid);
- 2) languse perioodide hulgast eemaldatakse perioodid, kus sobitatud mudeli parameetrid on füüsiliselt ebatõenäolised või sobituse viga on lubatavatest veapiiridest suurem;
- 3) mudeli sobitatud kordajate hulgas on eksponentsiaalse languse matemaatiline piirväärtus b , mis on ühtlasi ka teoreetiliseks välisõhu CO₂-sisalduseks ehk baasnivooks.
- 4) Langusperioodide järgi arvutatud baastasemete perioode võrreldakse BLC algoritmi järgi tuvastatud baastasemega vastavalt ajalise ülekatte perioodidel.

Piirkontsentratsioonil põhinev meetod jäeti tööst välja järgnevatel põhjustel:

- meetodi rakendamise eeldustele vastavaid, häiringuteta langusperioode oli väga vähe, kuna
 - sageli toimus hoopis taseme kiire alanemine ventilatsiooni tööaegadel ning seega oli langusperiood kaetud vaid ca 3 andmepunktiga, mis on eksponentsiaalse kõvera usaldusväärseks sobitamiseks ebapiisav;

- langusperioodidel esines sageli häiringuid, sh anomaalseid hüppeid ja murdepunkte;
- meetod ei suutnud pakkuda täiendavat täpsust ega usaldusväärset baastaseme arvutamisel, kuna
 - piirkontsentratsiooni järgi leitud baastase oli äärmiselt tundlik ajaperioodi määratlusele;
 - eksponentsiaalsete languskõverate sobitamisel on piirväärtuse stabiilseks määramiseks vaja andmepunkte koonduvast alast, mis sageli puudus. Kui koonduv ala oli olemas, ühtis see BLC meetodil leitud alumise protsentiili väärtusega ja ei andnud valideerimiseks vajalikku lisandväärtust;
- meetod oli liialt arvutusmahukas, nõudis täpset languste tuvaastusperioodide algoritmi ning puudus robustsus.

Järgneval lehel on esitatud näidis piirkontsentratsioonide meetodil rakendamisest.



Joonis LISA 1 piirkontsentratsioonide meetodil sobitatud languskõverad (lilla joon), baastasemed (roheline katkendjoon) ning tuvastatud langusperioodid (punane ala). BLC algoritmiga tuvastatud baastase on märgitud helesinise joonega.