



TAMPEREEN TEKNILLINEN YLIOPISTO

MIKKO TUOHIMAA
MATLAB-POHJAINEN TYÖKALU SÄÄHAVAINTOJEN
ANALYSOINTIIN OLKILUODON ALUEELLA
Diplomityö

Tarkastaja:
professori Tarmo Lipping
Tarkastaja ja aihe hyväksytty
Talouden ja rakentamisen
tiedekuntaneuvoston kokouksessa
8. kesäkuuta 2016

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO

Johtamisen ja tietotekniikan koulutusohjelma

TUOHIMAA, MIKKO: MatLab-pohjainen työkalu säähavaintojen analysointiin

Olkiluodon alueella

Diplomityö, 50 sivua, 4 liitesivua

Lokakuu 2016

Pääaine: Signaalinkäsittely

Tarkastaja: professori Tarmo Lipping

Avainsanat: Tilastollinen analyysi, datan korjaaminen, vallitseva tuulen suunta, kasvukausi

Posiva on asiantuntijaorganisaatio, joka vastaa käytetyn ydinpolttoaineen loppusijoituksesta sekä siihen liittyvästä tutkimustoiminnasta Olkiluodon saarella. Yhtenä monista tutkimuskohteista on saaren ympäristö, ja tutkimuksen osana toteutetaan ympäristön monitorointiohjelmaa. Osana tätä ohjelmaa kerätään havaintoja saaren säätilasta. Havainnoista muodostuvaa aineistoa analysoimalla voidaan muodostaa kuva saaren sääolosuhteista. Posivalla on käytössä MatLab-ympäristöön toteutettu työkalu, jolla aineistoa käsitellään ja analysoidaan. Työkalusta on tämän diplomityön puitteissa toteutettu uudistettu versio.

Säähavaintoaineisto on usein virheellistä. Siitä puuttuu mittauksia tai osa mittauksista on joko liian suuria tai liian pieniä. Nämä virheelliset mittaukset aiheuttavat aukkoja aineistoon, joka on korjattava ennen kuin aineisto voidaan analysoida. Eri menetelmiin, joilla aineistoa korjataan, liittyy myös omat ongelmansa, joista keskeisin on menetelmän aiheuttama virhe. Hyvä korjausmenetelmä pyrkii minimoimaan tämän virheen.

Analysoitavia suureita ja laskettavia tunnuslukuja on useita, tunnuslukuja lasketaan myös useilta eri jaksoilta. Lisäksi mikäli aineistoa on jouduttu korjaamaan, sillä on vaikutusta analyysin luotettavuuteen: mitä enemmän aineistossa on ollut virheitä, sitä pienempi on luotettavuus. Analyysi on kuitenkin systemaattinen kokonaisuus, joka voidaan suorittaa tehokkaasti tarkoitusta varten suunnitellulla algoritmilla, joka laskee kaikki halutut tunnusluvut kaikilta jaksoilta ilman käyttäjän panosta. Analyysi laskee myös luotettavuuden kunkin tunnusluvun osalta.

Tämän diplomityön puitteissa on toteutettu ja testattu uudistettu versio työkalusta, johon toteutettujen korjausmenetelmien ja analyysin kuvaamiseen ja arvioimiseen kirjallinen osuus paneutuu. Toteutukseen liittyvät määritykset ja vaatimukset esitellään pikaisesti ja niiden täyttymistä myös arvioidaan. Työkalun uudistettua versiota on käytetty onnistuneesti säähavaintoaineiston analysointiin vuosilta 2013, 2014 ja 2015.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Management and Information Technology

TUOHIMAA, MIKKO: MatLab Based Tool for Analysing Weather Observations at Olkiluoto

Master of Science Thesis, 50 pages, 4 Appendix pages

October 2016

Major: Signal Processing

Examiner: Professor Tarmo Lipping

Keywords: Statistical analysis, data correction, prevailing wind direction, growth season

Posiva is an expert organisation which is responsible for the final disposal of spent nuclear fuel and the study of it at the island of Olkiluoto. One of the many subjects is the environment of the island, and as a part of that research Posiva carries out an environment monitoring programme. Part of the programme concerns observing weather of the island in a systematic fashion. In order to have an understanding on the climate conditions of the island the measurement data must be analysed properly. Handling and analysis of the data used to be performed with a MatLab based tool, and on Posiva's commission that tool has now been renewed in this Master's Thesis to better suit the needs of data handling and analysis.

Weather observation data are prone to errors such as missing measurements, or too low or too high values. These errors cause gaps in the measurement data which have to be repaired and validated before the analysis can start. The various methods proposed for data correction have their own issues as well, most notably the error each particular method causes when applied. A good method attempts to minimize the error.

There are plenty of quantities to be processed and key figures to be calculated, and from several different periods of time in the analysis of the observation data. Furthermore, if the data have to be repaired it will have an impact on the reliability of the results: the more repairing done, the less reliable the results. Analysis is, however, a systematic entity, which can be performed through an algorithm specifically designed for the purpose, and which calculates all the applicable key figures for all the periods without any effort from the user. As a part of the analysis, an estimate of reliability is given for each key figure calculated.

A renewed version of the MatLab based tool has been implemented and tested in this Master's Thesis which focuses mainly on describing and evaluating the proposed data correction methods and the analysis of the measurement data. Requirements and specifications of the tool, and their fulfilment are only briefly discussed. The renewed version of the tool has been successfully used to perform the analysis of the weather observation data for years 2013, 2014 and 2015.

ALKUSANAT

Tämä diplomityö on tehty Tampereen teknillisen yliopiston Porin yksikössä Posivan toimeksiannosta. Työ on mahdollistanut tutustumisen Posivan tutkimustoimintaan sekä haasteisiin, joita sään mittaamiseen liittyy.

Kiitän professori Tarmo Lippingiä (TTY, Pori) ohjeista työn tekemiseen liittyen, työn tarkastamisesta, sekä kärsivällisyydestä, jota tämä minun puoleltani pitkäksi venähtänyt prosessi on varmasti vaatinut.

Lisäksi halua kiittää Tuomas Pereä (Posiva Oy) palautteesta, korjausehdotuksista sekä opastetusta kierroksesta Olkiluodon ydinvoimalan alueella. Kiitän myös Ari Ikosta (EnviroCase) lähdemateriaalista, ohjeistuksesta sekä itse toimeksiannosta, sekä Lasse Aroa (Luonnonvarakeskus) lähdemateriaalista. Kiitos myös Taavet Lippingille (TTY, Pori) testaamisesta, Juho Kuusistolle (Posiva Oy) kommentaareista ja testaamisesta, sekä Lauri Parviaiselle (Posiva Oy) kommentaareista.

23. lokakuuta 2016

Mikko Tuohimaa

SISÄLLYS

1	JOHDANTO.....	1
1.1	MetEnv -työkalun käyttötapaus ja vaatimukset.....	2
2	SÄÄHAVAINTOASEMAT JA -AINEISTO.....	5
2.1	Säähavaintoasemat.....	6
2.1.1	TVO:n havaintoasema.....	6
2.1.2	Metsäasemat.....	7
2.2	Suure, mittadata ja pätevät mittaukset.....	8
3	VALIDOINTI JA KORJAUS.....	12
3.1	Validointi.....	12
3.2	Korjausmenetelmät.....	14
3.2.1	Manuaalinen korjaus.....	16
3.2.2	Vakiolla täyttö.....	16
3.2.3	Lineaarinen interpolointi.....	17
3.2.4	Profilointi.....	18
3.3	Korjausmenetelmien arviointi.....	21
3.3.1	Testit.....	23
3.3.2	Yhteenveto korjausmenetelmistä.....	24
4	ANALYYSI.....	26
4.1	Laskentakanta ja analysoitavat jaksot.....	26
4.1.1	Analysoitavat jaksot.....	27
4.2	Analyysi.....	29
4.2.1	Perusanalyysi.....	29
4.2.2	Mittadatan eheys.....	32
4.2.3	Toisen vaiheen analyysit.....	34
4.2.4	Pitkän aikavälin analyysi.....	37
4.3	Vallitseva tuulen suunta.....	39
4.4	Kasvukausi.....	43
5	YHTEENVETO.....	47
	LÄHTEET.....	49
	LIITE A – OL-WOM2 -ASEMALTA MITATTAVAT SUUREET.....	51
	LIITE B – MITTADATATIEDOSTOESIMERKIT.....	53

TERMIT JA NIIDEN MÄÄRITELMÄT

DI	Mittadatan eheys (<i>Data Integrity</i>) on aitojen mittausten määrän suhde odotettuun määrään joltakin jaksolta. Suhdeluku voidaan esittää myös prosentteina.
ICP ICP Forests	ICP Forests (<i>The International Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests</i>) on kansainvälinen ohjelma, joka tähtää kattavaan tietokantaan metsien terveydentilasta yleisesti Euroopassa sekä myös Suomessa.
Interpolointi	Eräs mittadatan korjausmenetelmä, jossa puuttuvista mittauksista johtuvat aukot mittadatassa korjataan lineaarisella interpoloinnilla.
Kasvukausi	Eräs analyyseistä, jossa keskilämpötilojen perusteella määritetään se jakso vuodesta, jolloin kasvit kasvavat.
Keskimääräinen erotus	Keskimääräinen erotus kertoo keskimääräisen muutoksen suuruuden lisäksi myös muutoksen suunnan.
LTA	Pitkän aikavälin analyysi (<i>Long Term Analysis</i>) on eräs mittadatalle suoritettava analyysi, jossa lasketaan pitkän aikavälin keskiarvoja.
LTAM	Pitkän aikavälin kuukausittainen analyysi (<i>Long Term Analysis, Monthly</i>) on eräs mittadatalle suoritettava analyysi, jossa lasketaan pitkän aikavälin kuukausittaisia keski- ja ääriarvoja.
MatLab	MatLab (<i>Matrix Laboratory</i>) on MathWorks -yhtiön laskentaan tarkoitettu ohjelmisto ja siinä käytettävä ohjelmointikieli.
MetEnv	MetEnv on MatLabilla toteutettu työkalu, jolla automatisoidaan mittadatan käsittelyä ja analysointia.
MSE	Keskineliövirhe (<i>Mean Squared Error</i>) on luku, joka kertoo virheen suuruuden aitojen ja estimoitujen mittausten välillä.
NaN	Liukuluvuilla esitettävä ei-numeerinen luku (<i>Not-a-Number</i>), jonka arvo on määrittelemätön.
OL-WOM	Olkiluodon saarella oleva säähavaintoasema tai mitta-asema (<i>Olkiluoto Weather Observation Mast</i>), jossa mitataan useita eri suureita toisinaan myös useammalta eri korkeudelta, jolloin asemaan kuuluu myös masto.
PAR	Fotosynteettisesti aktiivinen säteily (<i>Photosynthetically Active Radiation</i>) on eräs mitattavista suureista.
Posiva	Posiva on organisaatio, joka vastaa sekä Olkiluodon että Loviisan ydinvoimaloiden tuottaman ydinjätteen loppusijoituksesta.

Profilointi	Eräs mittadatan korjausmenetelmistä, jossa puuttuvista mittauksista johtuvat aukot mittadatassa pyritään korjaamaan referenssidatalla.
PWD	Vallitseva tuulen suunta (<i>Prevailing Wind Direction</i>) on eräs mittadatalle suoritettavista analyyseistä. Se ilmaisee yleisimmän tuulen suunnan joltakin jaksolta ja sen määrittämiseksi on kaksi menetelmää.
PWDA	Vallitseva tuulen suunta kertymän perusteella (<i>Prevailing Wind Direction by Accumulation</i>) määrittää yleisimmän tuulen suunnan joltakin jaksolta huomioiden tuulen suunnan lisäksi myös tuulen nopeuden.
PWDF	Vallitseva tuulen suunta frekvenssin perusteella (<i>Prevailing Wind Direction by Frequency</i>) määrittää yleisimmän tuulen suunnan joltakin jaksolta huomioiden vain tuulen suunnan.
STUK	Säteilyturvakeskus on säteily- ja ydinturvallisuutta valvova viranomainen Suomessa.
Tehoisa lämpötila	Kasvukauteen läheisesti liittyvä tunnusluku, joka lasketaan päivittäisten keskilämpötilojen perusteella.
Terminen kasvukausi	Ks. kasvukausi
TVO	Teollisuuden Voima Oyj on Olkiluodon ydinvoimalan omistaja.
WebMET	<i>The Meteorological Resource Center</i> on verkkosivusto, joka tarjoaa mm. tietoa liittyen sääilmiöihin sekä säädatan käsittelyyn.
WOM	Ks. OL-WOM

1 JOHDANTO

Posiva on organisaatio, joka vastaa sekä Olkiluodon että Loviisan ydinvoimaloiden tuottaman käytetyn ydinpolttoaineen loppusijoituksesta. Käytetty polttoaine on tarkoitettu loppusijoittamaan syväälle Olkiluodon kallioperään. Olkiluoto on saari, joka kuuluu Eurajoen kuntaan, ja se sijaitsee Selkämeren rannikolla Etelä-Satakunnassa. Saarella sijaitsevat myös Teollisuuden Voima Oyj:n (TVO) omistamat ydinlaitosyksiköt Olkiluoto 1 ja Olkiluoto 2, sekä rakenteilla oleva yksikkö Olkiluoto 3.

Posiva tutkii käytetyn ydinpolttoaineen loppusijoitusta Olkiluodon saarella. Esimerkiksi kallioperän tutkimusta varten saarelle on muun muassa louhittu tutkimusluolasto ONKALO. Yhtenä monista tutkimuskohteista on myös saaren ympäristö, jonka tutkimus toteutetaan pääasiassa osana ympäristön monitorointiohjelmia. Osana tätä ohjelmaa kerätään havaintoja Olkiluodon saaren säätilasta. Näistä havainnoista muodostuvaa säähavaintoaineistoa analysoimalla voidaan muodostaa kuva saaren sääolosuhteista.

Säähavaintoaineistoja hyödynnetään biosfäärimallinnuksessa hydrologisen mallinnuksen kautta. Niitä tarvitaan estimoimaan ydinlaitosten päästöjen leviämistä ja laskeumaa, ja ne ovat myös avuksi selittämään toimialan yleisiä ympäristövaikutuksia. Näiden lisäksi säätilaa mitataan osana lakisäätteistä seuranta- ja mahdollisiin hätätiloihin varautumista varten. Tällä diplomityöllä ei ole kuitenkaan mitään tekemistä varsinaisen ydinturvallisuuden kanssa, vaan työssä keskitytään vain säähavaintoaineistojen käsittelyyn ja analysointiin. [1][2]

Havaintoaineistoja on kahdelaisia ja ne eroavat toisistaan tavasta, jolla niitä kerätään. Esimerkiksi lumihavaintoja kerätään manuaalisesti kiertämällä mitta-asetat ja lukemalla mittarit jokseenkin säännöllisin välein. Sen sijaan säähavaintoja tehdään tätä säännöllisemmin ja yhtäjaksoisemmin. Tällaiset mittaukset suoritetaan mitta-asetalle asennetuilla antureilla ja automaattisella keräyslaitteistolla. Näitä automatisoituja säähavaintoasemia on Olkiluodon saarella kaikkiaan viisi, joista jokainen mittaa joukon erilaisia suureita toisinaan myös usealta eri korkeudelta. Mitattavia suureita ovat muun muassa ilman lämpötila, ilman kosteus, sademäärä ja maaperän lämpötila.

Säähavaintoasemien tuottama aineisto analysoidaan vuosittain. Aineistosta lasketaan useita tunnuslukuja, kuten keskiarvoja, kertymiä ja ääriarvoja vuorokausien, kuukausien ja vuosien jaksoilta. Näiden lisäksi tunnuslukuja lasketaan myös pitkältä aikaväliltä kattaen koko havaintohistorian. Aineiston, jota on vähintään 30:lta vuodelta, perusteella voidaan jo tehdä havaintoja ilmastosta [17].

Posivalla on käytössä MatLab-ympäristöön aiemmin toteutettu työkalu, jonka tarkoitus on ollut helpottaa mittadatan korjaamista ja automatisoida havaintoaineiston analysointi. Työkalu kulkee nimellä MetEnv. Sen käyttö on kuitenkin ollut kankeaa ja toteutuksessa on havaittu sekä puutteita että suoranaisia virheitä. Tämän työn tarkoitus on toteuttaa uusi versio MetEnv-työkalusta, jonka käytettävyys on parempi ja toiminta tarkoituksenmukaisempaa sekä tietysti virheetöntä. MetEnvin uusi versio toteutetaan edelleen MatLabille.

Tässä diplomityössä keskitytään pääasiassa säähavaintoasemien tuottaman havaintoaineiston korjaamiseksi toteutettujen menetelmien yksityiskohtaiseen kuvaamiseen ja arviointiin, sekä aineistolle suoritettavan analyysin yksityiskohtaiseen kuvaamiseen. Näin ollen tämä diplomityö myös dokumentoi MetEnvin uuden version keskeisen toiminnallisuuden.

Luku 2 käsittelee säähavaintoasemia, mitattavia suureita sekä asemien tuottamaa havaintoaineistoa. Luvussa määritellään mitä suureella tämän työn kontekstissa täsmälleen tarkoitetaan. Luku liitteineen listaa kaikki suureet, joita mitataan, sekä mittauskorkeudet, ja esittelee lyhyesti havaintoaineiston formaattia. Näiden lisäksi esitellään aineistoon kiinteästi liittyvä statusluokitus, joka kertoo, miten aineistoa on mahdollisesti käsitelty. Statusluokitus on tärkeä elementti arvioitaessa aineistolle suoritettavan analyysin luotettavuutta.

Luku 3 keskittyy säähavaintoaineiston validointiin ja korjaamiseen. Aineisto tarkistetaan puuttuvien, liian suurien ja liian pienien arvojen varalta. Jos näitä puutteita havaitaan, aineisto ei ole validi ja se on näin ollen korjattava. Luvussa kuvataan manuaalisen korjaamisen lisäksi kolme automaattista menetelmää, jotka korjaavat koko vuoden aineiston kerralla. Nämä menetelmät ovat vakiolla täyttö, interpolointi ja profilointi. Lopuksi luvussa arvioidaan näitä menetelmiä niiden aiheuttaman virheen perusteella, sekä arvioidaan menetelmien soveltuvuutta eri suureille ja eri tilanteisiin.

Luku 4 kuvaa analyysin yksityiskohtaisesti. Analyysi on jaettu neljään osaan: perusanalyysi, toisen vaiheen analyysi, pitkän aikavälin analyysi sekä kasvukauden määrittäminen. Perusanalyysissä validista säähavaintoaineistosta lasketaan keskiarvoja, kertymiä ja ääriarvoja päivittäisille, kuukausittaisille ja vuotuisille jaksoille. Tämän lisäksi määritellään myös mittadatan eheys, joka vaikuttaa suoraan analyysin luotettavuuteen. Perusanalyysin tuloksia käytetään sitten jatkoanalyysiin, eli toisen vaiheen analyysiin, pitkän aikavälin analyysiin ja kasvukauden määrittämiseen. Luvussa käsitellään myös ongelmia, joita kasvukauden määrittämiseen liittyy. Viimeisessä luvussa, eli luvussa 5 on vielä yhteenveto.

1.1 MetEnv -työkalun käyttötapaus ja vaatimukset

Kuten edellä todettiin, MetEnv on MatLab-ympäristöön MatLab-ohjelmointikielellä toteutettu työkalu, jonka tarkoituksena on ollut helpottaa säähavaintoaineiston

korjaamista ja automatisoida sen analysointi. Työkalussa on graafinen käyttöliittymä ja sen on ollut tarkoitus käsitellä myös havaintoaineistoa, jota kerätään manuaalisesti ja epäsäännöllisemmin, ja joka siten poikkeaa muusta aineistosta. Työkaluun on toteutettu myös erilaisia interpolointimenetelmiä puuttuvista havainnoista aiheutuneiden aukkojen paikkaamiseksi.

Työkalun uusi versio on täysin uusi toteutus, jossa ei ole käytetty vanhaa lähdekoodia muutoin kuin korkeintaan spesifikaationa eräiden yksittäisten parametrien osalta. Uusi versio on toteutettu edelleen MatLab-ympäristöön MatLab-kielellä ja siinäkin on graafinen käyttöliittymä. Toteutustyön painopiste on ollut käyttöliittymässä, jolla havaintoaineiston käsittely ja korjaaminen olisi mahdollisimman vaivatonta, ja aineiston automaattisessa analyysissä, joka tuottaisi kaiken tarvittavan datan ympäristön monitorointiraportteja (esimerkiksi *Results of Monitoring at Olkiluoto in 2013 – Environment*) silmällä pitäen.

MetEnv-työkalun uuden version käyttötapaus on kolmivaiheinen:

1. Käyttäjällä on saatavilla säähavaintoaineisto viimeisimmältä mittausvuodelta, jonka hän haluaa tuoda työkaluun analysoitavaksi. Aineisto koostuu useasta tekstitiedostosta, jotka noudattavat ennalta määriteltyä formaattia. Nämä tiedostot sisältävät mittauksia, joita on kerätty vuoden jokaiselta tunnilta usealle eri suureelle. Käyttäjän ei itse tarvitse tuottaa tätä aineistoa tai osallistua edes sen mahdolliseen esikäsittelyyn.
2. Havaintoaineiston odotetaan olevan säännöllistä ja sisältävän pätevän mittauksen vuoden jokaiselta tunnilta. On kuitenkin tavallista, että osa mittauksista on selvästi virheellistä tai uupuu kokonaan. Jotta käyttäjä voi aloittaa analyysin, pitää mahdollisesti virheellinen ja puutteellinen aineisto ensin korjata tarpeen niin vaatiessa. Koska on mahdollista, että mittauksia esimerkiksi puuttuu paljon, ei ole ollenkaan mielekästä, että käyttäjän tulisi paikata aukkoja manuaalisesti vaan hänellä on oltava muitakin menetelmiä, jotka ovat tehokkaampia sekä vähemmän virhealttiimpia käyttäjän tekemille virheille. Kun käyttäjä on korjannut aineiston, MetEnv toteaa aineiston olevan korjattu ja tallentaa sen paikalliseen tietokantaan, joka on tiedosto.
3. Käyttäjä aloittaa aineiston analyysin. Hän haluaa aineiston analysoitavan ensisijaisesti silmällä pitäen vuotuisia monitorointiraportteja. Koska vaihe on täysin systemaattinen, hän ei itse osallistu analyysiin millään tavalla vaan poimii vain tulokset raporttiin analyysin valmistuttua.

Tästä käyttötapauksesta ja monitorointiraporteista sekä Posivan asiantuntijoiden tahoilta tulevat seuraavat vaatimukset MetEnv-työkalulle:

1. Työkaluun tulee voida tuoda määrämuotoinen aineisto.
2. Aineistoa tulee käsitellä mahdollisimman yleisellä tasolla välttämättä suurekohtaisia sidoksia.
3. Aineistosta tulee tunnistaa ja merkitä puuttuvat mittaukset.

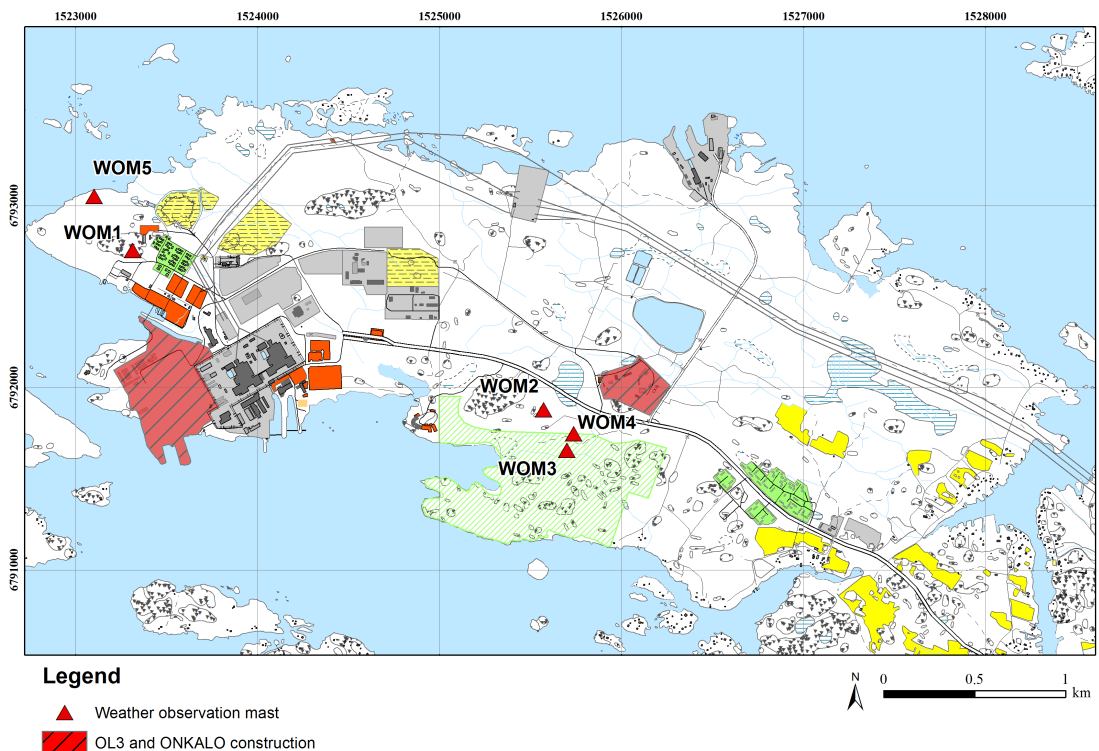
4. Aineistosta tulee tunnistaa, korjata ja merkitä liian suuret ja liian pienet arvot.
5. On toteutettava menetelmiä, joilla paikataan puuttuvista mittauksista johtuvia aukkoja aineistossa. Näitä ovat ainakin lineaarinen interpolointi ja korvaaminen samankaltaisella, saatavissa olevalla aineistolla (profilointi).
6. Korjausmenetelmiä tulee voida suorittaa koko vuoden aineistolle siten, että jokaista aukkoa ei tarvitse korjata yksitellen.
7. Aidot ja korjatut mittaukset tulee voida erottaa toisistaan selvästi.
8. Aitojen mittausten määrä tulee olla saatavilla.
9. Analyysissä lasketaan keskiarvoja, kertymiä ja ääriarvoja päivittäisille, kuukausittaisille ja vuotuisille jaksoille koko aineistosta.
10. Analyysissä lasketaan kuinka suuri osuus kuhunkin analyysiin huomioiduista mittauksista on aitoja mittauksia.
11. Analyysissä lasketaan myös tunnuslukuja päivittäisen analyysin tuloksista, pitkän aikavälin analyysi, sekä määritetään kasvukausi.
12. Analyysin tulokset on helposti poimittavissa monitorointiraporttiin, sekä raportissa esiintyviä kuvaajia varten.

Luvuissa 2 ja 3 selvitetään miten vaatimukset 3 – 8 on huomioitu työkalun uuden version toteutuksessa, kun taas luku 4 keskittyy vaatimukseen 9 – 12. Tässä työssä ei kuitenkaan keskitytä varsinaiseen ohjelmointityöhön tai -menetelmiin, jotka liittyvät esimerkiksi tiedostojen lukuun. Tämän diplomityön fokus on korjausmenetelmien ja analyysin yksityiskohtaisessa ja tarkassa, matemaattisessa kuvaamisessa. Lopuksi, kaikki viittaukset MetEnv-työkaluun viittaavat jatkossa aina työkalun uuteen versioon ellei erikseen mainita.

2 SÄÄHAVAINTOASEMAT JA -AINEISTO

Olkiluodon saaren ympäristö on yksi Posivan monista tutkimuskohteista, ja osana tutkimusta toteutetaan ympäristön monitorointiohjelmaa, jossa muun muassa kerätään havaintoja Olkiluodon saaren säätilasta. Näistä havainnoista muodostuvaa säähavaintoaineistoa sitten analysoidaan, jotta voidaan muodostaa kuva saaren sääolosuhteista. Analyysin tulokset julkaistaan vuosittain ympäristön monitorointiraporteissa. [4]

Tässä luvussa käsitellään säähavaintoasemia, suureita ja havaintoaineistoa. Suureita on paljon ja jokaista mitataan säännöllisin väliajoin, mikä muodostaa yhtäjaksoisen havaintoaineiston. Olkiluodon saarella on yhteensä viisi automaattista säähavaintoasemaa, jotka tuottavat tällaista säännöllistä ja yhtäjaksoista havaintoaineistoa, jota kutsutaan jatkossa mittadataksi. Näiden asemien sijainnit näkyvät kuvassa 1.



Kuva 1: Säähavaintoasemat (Weather Observation Mast – WOM) Olkiluodon saarella, jotka mittaavat suureita automaattisesti. Kartan sommittelu: Tuomas Pere/Posiva Oy.

2.1 Säähavaintoasemat

Olkiluodon saarella on viisi automatisoitua säähavaintoasemaa. Ne mittaavat useita suureita säännöllisin välein, yhtäjaksoisesti ja pääasiassa maan tasolta. (Maatason mittauskorkeus on itse asiassa 2 metriä, eikä suinkaan maan pinta.) Kahdella asemista on masto, joista ensimmäinen mittaa suureita jopa 100 metrin ja toinen 24 metrin korkeudelta. Asemat (tai mastot) on nimetty seuraavan säännön mukaan: OL-WOM n (*Olkiluoto Weather Observation Mast*), missä n saa siis arvot 1, 2, 3, 4 tai 5. Myös lyhenteellä WOM voidaan viitata näihin havaintoasemiin.

2.1.1 TVO:n havaintoasema

OL-WOM1 on TVO:n havaintoasema ja se sijaitsee saaren länsipäädyssä (ks. kuva 1). Sen tarkoitus ja toiminta perustuu säteilyturvakokeskuksen (STUK) ohjeeseen [7], joka puolestaan pohjaa ydinenergialakiin. STUK ohjeistaa mm. seuraavaa:

”Jotta radioaktiivisten aineiden päästöjen leviämisen ja väestön säteilyannosten arviointi on mahdollista, on ydinvoimalaitoksen meteorologisten mittausten avulla saatava tiedot ainakin tuulen suunnasta ja nopeudesta, ilmakehän rajakerroksen stabiiliudesta sekä sateesta.”

”Maanpinnan läheltä on mitattava ilman lämpötila, ilmanpaine, ilman kosteus sekä sademäärä ja -aika.”

”Säämastolla alimmat mastomittaukset on tehtävä 2–3 kertaa korkeammalla kuin lähimaaston rosoisuuselementtien (kuten puiden) keskimääräinen korkeus. Ylimmän mittauskorkeuden on vastattava vähintään ydinvoimalaitoksen poistoilmapiipun korkeutta.”

”Säämastolla lämpötila- ja tuulimittausten mittausantureita on sijoitettava edellä mainittujen korkeuksien lisäksi vähintään yhdelle näiden väliselle mittauskorkeudelle.”

Taulukkoon 1 on koottu OL-WOM1 -asemalta mitattavat suureet yksiköineen [6]. Tuulen suunnan sijaan mitataan leviämissuuntaa, joka on vastakkainen tuulen suunnalle. Lämpötilaa mitataan kaikkiaan neljältä eri korkeudelta sekä tuulen nopeutta ja suuntaa kolmelta eri korkeudelta. OL-WOM1 -asemalta on saatavilla mittadataa yhtäjaksoisesti aina tunnin välein lähtien vuodesta 1992.

Tämä TVO:n säähavaintoasema liittyy ensisijaisesti ydinturvallisuuteen. TVO toimittaa mittadatan Posivalle, joka puolestaan hyödyntää sitä ympäristön tutkimuksessa. Tällä diplomityöllä ei näin ollen ole mitään tekemistä TVO:n toiminnan saati ydinturvallisuuden kanssa.

Taulukko 1: OL-WOM1 -aseman suureet, yksiköt ja mittauskorkeudet

Suure [yksikkö]	Mittauskorkeus [m]			
	2	20	60	100
Lämpötila [°C]	X	X	X	X
Tuulen nopeus [m/s]		X	X	X
Leviämissuunta [°]		X	X	X
Leviämissuunnan hajonta [°]		X	X	X
Ilmanpaine [hPa]	X			
Ilman suhteellinen kosteus [%]	X			
Sademäärä [mm]	X			
Sadeaika [min]	X			

2.1.2 Metsäasemat

Säähavaintoasemat 2 – 5 toimivat osana metsien intensiivitutkimusta. Asemat 2, 3 ja 4 sijaitsevat lähellä toisiaan Olkiluodon saaren eteläisellä puolella, ja asema 5 on aivan saaren länsipäässä lähellä meren rantaa (ks. kuva 1).

Intensiivitutkimus tarkoittaa käytännössä automatisoitua, taajempaa ja säännöllisesti toistuvaa mittaamista verrattuna ekstensiivitutkimukseen, jossa mittauksia tehdään manuaalisesti, harvemmin ja epäsäännöllisemmin. Nämä asemat ovat samankaltaisia ICP Forests -ohjelmassa määriteltyjen intensiiviasemien kanssa [3, s. 7]. ICP Forests (*International Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests*) on kansainvälinen ohjelma, joka tähtää kattavaan tietokantaan metsien terveydentilasta yleisesti Euroopassa [8] sekä myös Suomessa [9].

Mittaukset suoritetaan pääasiassa maan tasolla (2 m), mutta OL-WOM2 -asemalla on vielä masto tuuli-, sekä latvuksen ylä- ja alapuolisia mittauksia varten [1, s. 123]. Taulukossa 2 esitetään mittatavat suureet sekä käytettyjen anturien lukumäärät [6][9]. Näiden lisäksi OL-WOM2 -asemalta mitataan vielä joukko muita suureita. Täydellinen lista niistä löytyy liitteestä A.

Suureita mitataan 10 minuutin intervallein ja mittauksista lasketut keski- sekä ääriarvot (minimit ja maksimit) kirjataan tunnin välein [1, s. 123][10]. Näin muodostetussa mittatietokannassa on siis arvoja yhtäjaksoisesti aina tunnin välein, jos vain data on ehjää.

Kuten taulukoista 1 ja 2 sekä varsinkin liitteestä A käy selväksi, mitattavia suureita on paljon. Määritellyt analyysit (luku 4) suoritetaan toki kaikille suureille, mutta ainoastaan osa tuloksista raportoidaan Posivan seurantaraportissa [4][5].

Taulukko 2: Metsäasemien suureet yksiköineen ja anturien lukumäärät

Suure [yksikkö]	OL-WOM			
	2	3	4	5
Ilman lämpötila [°C]	1	1	1	1
Ilman suhteellinen kosteus [%]	1	1	1	1
Ilmanpaine [hPa]	1	0	0	0
Sademäärä [mm]	1	1	1	1
Maaperän lämpötila (−10 cm) [°C]	3	3	3	3
Maaperän lämpötila (−20 cm) [°C]	3	3	3	3
Maaperän lämpötila (−30 cm) [°C]	1	1	1	1
Maaperän lämpötila (−40 cm) [°C]	1	1	1	1
Maaperän lämpötila (−50 cm) [°C]	1	1	1	1
Maaperän lämpötila (−60 cm) [°C]	1	1	1	1
Maaperän lämpötila (−70 cm) [°C]	1	1	1	1
Maaperän lämpötila (−80 cm) [°C]	1	1	1	1
Maaperän lämpötila (−90 cm) [°C]	1	1	1	1
Maaperän kosteus (−20 cm) [%]	2	2	3	2

2.2 Suure, mittadata ja pätevät mittaukset

Suureista saadaan mittauksia siis yhtäjaksoisesti tunnin välein. Mittausten muodostama joukko on järjestetty joukko, jossa yksittäinen alkio voi toistua. Tätä joukkoa voi ajatella vektorina, aikadiskreettinä signaalina tai funktiona, jonka määrittelyalue on luonnolliset luvut ja arvoalue reaalityyliset luvut. Koska mittaukset ovat desimaalilukuja esimerkiksi sadasosien tarkkuudella, arvoalue on käytännössä rationaaliluvut.

Tässä työssä sekä MetEnv-työkalussa tätä joukkoa kutsutaan mittadatakseksi ja sitä käsitellään M -ulotteisena reaaliavaruuden vektorina. Olkoon \mathbf{x} jokin tällainen mittadatavektori (tai datavektori):

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix} \in \mathbb{R}^M \iff x_i \in \mathbb{R}, \quad i = 1, 2, \dots, M. \quad (1)$$

Mittadata on rajoitettua. On olemassa reaaliavaruuden $H < \infty$, $L > -\infty$ ja $L < H$ siten, että

$$L \leq x_i \leq H, \quad i = 1, 2, \dots, M. \quad (2)$$

Mittadataa odotetaan olevan yhtäjaksoisesti tunnin välein, ja sitä käsitellään tavallisesti koko vuosi kerralla. Vuodessa on 8760 tuntia ja karkausvuodessa 8784 tuntia. Vuoden ensimmäinen mittaus on aina ajanhetkeltä 1.1. kello 0:00, ja viimeinen 31.12. kello 23:00 OL-WOM1 -mittauksissa. Muissa asemissa mittaukset ovat myöhässä tunnilla, eli ensimmäinen mittaus on 1.1. klo. 1:00 ja viimeinen seuraavana vuonna 1.1. klo. 0:00 [10][11]. Näin ollen datavektorin (1) indeksi i on itse asiassa tunnin järjestysnumero alkaen jostakin tietystä ajankohdasta, kuten vaikkapa 1.1.2013 klo. 0:00 (jolloin datavektori olisi OL-WOM1 -asemalta).

Mittadata tuodaan tekstitiedostoista, jotka ovat määrämuotoisia. OL-WOM1 -tiedostoissa dataa on aina yhden kuukauden ajalta ja kaikilta antureilta, kun taas asemien 2 – 5 tiedostot sisältävät datan koko vuoden ajalta mutta ainoastaan yhdeltä anturilta. Se, miten data on tiedostoihin jaoteltu ja mikä tiedostojen eksakti muoto on, ei ole oleellista. Sen sijaan oleellista on, että tiedostot ovat säännönmukaisia, koska toisinaan niistä puuttuu mittauksia.

Mittaukset ovat tiedostoissa riveittäin siten, että jokaisen rivin alussa on aikaleima, jota seuraa yksi tai useampi mittaus (mitta-asemasta riippuen). Mittadat tiedostojen muoto on esitetty esimerkein liitteessä B. Aikaleimasta selviää päivämäärä ja tunti, jolta kyseinen mittaus on. Kuitenkin toisinaan riveiltä puuttuu mittauksia, tai rivejä uupuu kokonaan. Tiedostojen säännönmukaisuutta hyödyntäen näistä tilanteista voidaan selvittää, ja tämä onkin yksi tämän työn keskeisimmistä tavoitteista.

Puuttuvat mittaukset tai rivit aiheuttavat aukkoja mittadataan, joka on jotenkin paikattava. Luvussa 3 esitellään menetelmiä, joilla paikata dataa. Sitä ennen on kuitenkin varmistuttava siitä, että itse datavektori on täsmälleen odotetun pituinen, ja että aukkopaidat ovat tiedossa. Vektorin odotettu pituus ei ole hankala päätellä, koska tiedämme miltä vuodelta mittaukset ovat, ja että mittauksia on aina tunnin välein. Aukot merkitään mittadataan NaN-arvoina (*Not-a-Number*), jotka eivät siis ole numeerisia (vaikka arvo esitetäänkin liukuluvuilla) vaan määrittelemättömiä arvoja.

Puuttuvista tai pätemättömistä mittauksista pidetään kirjaa kuten ICP ehdottaa [10]. Jokaiselle mittatavektorin alkionle annetaan taulukossa 3 esitettyjen luokkien mukainen status, jotka kootaan omaan vektoriin. Tämä statusvektori on edelleen M -ulotteinen vektori, jonka alkiole kuuluvat nyt statusluokkien joukkoon. Olkoon S statusluokkien joukko ja s jokin statusvektori. Näin ollen, sekä taulukon 3 mukaan:

$$S = \{0, 1, 2, 3, 4, 7, 8, 9\} \text{ ja} \quad (3)$$

$$\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{pmatrix} \in S^M \iff s_i \in S, \quad i = 1, 2, \dots, M. \quad (4)$$

Statusvektori on synkronoitu datavektoriin, toisin sanoen vektorit ovat yhtä pitkät ja datavektorin alkia x_i vastaava statusluokitus on s_i .

Taulukko 3: Statusluokat ja niiden merkitys. Luokkia puuttuu välistä, koska työkalun kehityksen yhteydessä ilmeni tarvetta luokituksille, jotka kuitenkin siivoutuivat pois lopullisesta versiosta.

Status	Selitys
0	Aito, ja siten pätevä mittaus, ts. mittaus on tiedostossa ja se on määriteltujen raja-arvojen sisällä.
1	NaN-arvo. Mittaus ei ole tiedostossa (tai sitä ei voi tulkita numeerisena) ja ei ole täten pätevä.
2	Alkuperäinen mittaus alittaa alarajan. Mittaus on korvattu alarajalla ja sitä voidaan pitää pätevänä.
3	Alkuperäinen mittaus ylittää ylärajan. Mittaus on korvattu ylärajalla ja sitä voidaan pitää pätevänä.
4	Mittaus on korjattu manuaalisesti ja sitä voidaan pitää pätevänä.
7	Puuttuva mittaus on korjattu interpoloimalla (luku 3) ja sitä voidaan pitää pätevänä.
8	Puuttuva mittaus on korvattu vakiolla (luku 3) ja sitä voidaan pitää pätevänä.
9	Puuttuva mittaus on korjattu profiloinnilla (luku 3) ja sitä voidaan pitää pätevänä.

Suurin osa mittauksista on päteviä ja niiden statusluokka on siten 0. Vain pätevät mittaukset otetaan huomioon dataa analysoitaessa, joten mittauksille, joiden status on 1, tulee suorittaa korjaavia toimenpiteitä. MetEnv ei edes anna käyttäjän aloittaa analyysiä ennen kuin kaikki aukot on paikattu. Näiden toimenpiteiden seurauksena mittaus saa sitten luokituksen 4, 7, 8 tai 9 suoritetusta operaatiosta riippuen. Arvoalue tarkistetaan sekä korjataan automaattisesti mittadatan tuonnin yhteydessä. Nämä korjaavat toimenpiteet käsitellään luvussa 3.

Taulukosta 3 voidaan poimia vain osa statusluokista, kun halutaan kaventaa päteviksi katsottujen mittausten joukkoa kulloiseenkin tilanteeseen soveltuvaksi. Olkoon V (tilanteeseen) pätevien statusluokkien joukko, jolle pätee:

$$V \subset S \quad \wedge \quad 1 \notin V. \quad (5)$$

V on siis aito osajoukko, koska NaN-arvoja ei koskaan pidetä pätevinä.

Määritellään myös funktio $v: S \rightarrow \{0, 1\} \subset \mathbb{R}$, joka huomioi vain pätevien mittausten vaikutuksen:

$$v(s; V) = \begin{cases} 1 & , s \in V \\ 0 & , s \notin V \end{cases}. \quad (6)$$

Funktion argumentti on statusluokka ja sille annetaan parametrina määritelmän (5) mukainen pätevien statusluokkien joukko.

Määritellään vielä yksi tärkeä tunnusluku, jota tarvitaan mittadatan eheyden ja siten analyysien luotettavuuden arvioinnissa. Data- ja siten myös statusvektorin pituus on M . Olkoon \mathbf{s} edelleen mittadatavektoria \mathbf{x} vastaava statusvektori. Pätevien mittausten lukumäärä $N \leq M$ saadaan nyt funktion (6) avulla:

$$N = \sum_{i=1}^M v(s_i; V). \quad (7)$$

3 VALIDOINTI JA KORJAUS

Edellisessä kappaleessa esiteltyjen mitta-asemien tuottama data voi sisältää (ja usein sisältääkin) puutteellisia, epäilyttäviä tai virheellisiä arvoja. Tämä voi johtua vaikkapa vikaantuneista mittalaitteista, kalibroimattomista antureista tai säätilasta: anturi tai laite on voinut esimerkiksi yksinkertaisesti jäätyä. On tietysti myös mahdollista, että epäilyttävä arvo osoittautuukin lopulta päteväksi havainnoksi.

Validointi on prosessi, joka pyrkii havaitsemaan potentiaaliset sekä ilmiselvät virheet mittatiedossa. Kun virheelliset mittaukset ovat tiedossa, niitä voidaan korjata suureeseen parhaiten soveltuvalla menetelmällä. Tämän vaiheen tuloksena on validi mittadata, josta virheet on korjattu, ja joka on valmis analysoitavaksi. Validointia sekä korjausmenetelmiä käsitellään tässä luvussa. Menetelmien hyvyttä myös arvioidaan MatLabilla suoritettujen testien perusteella.

3.1 Validointi

Validoinnin tavoite on tunnistaa virheelliset tai puuttuvat arvot. Tavallinen virhetyyppi, jota mittadata sisältää on puuttuvat mittaukset. Tätä esiintyy esimerkiksi OL-WOM1-mittatiedossa vuodelta 2013 ilman suhteellisen kosteuden ja sademäärän yhteydessä. ICP määrittelee kaksi tarkistusta datan eheyden evaluoimiseksi: tarkistetaan, että mikään mitattavista suureista ei puutu ja mittadata on yhtäjaksoista [10].

Ensimmäistä tarkistusta ei erikseen tehdä, on MetEnvin käyttäjän vastuulla tuoda työkaluun kaikki vaaditut mittatiedostot. Sen sijaan yhtäjaksoisuus tarkistetaan automaattisesti mittatiedon tuonnin yhteydessä. Olkoon x ja s määritelmien (1) ja (4) mukaiset mittadata ja sitä vastaava statusvektori. Kuten luvussa 2 kerrottiin, kaikki mittatiedossa esiintyvät ei-reaaliset arvot (NaN-arvot) tulkitaan puuttuviksi, eli aukko-paikoiksi mittatiedossa. Statusvektoriin aukot merkitään seuraavasti:

$$s_i = \begin{cases} 1 & , x_i \notin \mathbb{R} \\ s_i & , x_i \in \mathbb{R} \end{cases} , i = 1, 2, \dots, M . \quad (8)$$

Status ei siis päivitetä, jos vastaava mittaus on reaalinen.

Toinen tavallinen virhetyyppi on selvästi liian suuret tai pienet arvot. Nämä on helppo havaita automaattisesti. Hankalampi on sen sijaan havaita mittauksia, jotka ovat epäilyttävän suuria tai pieniä, muttei kuitenkaan ylitä mitään asetettuja raja-arvoja. ICP [10][11] jatkaa määrittelemällä seuraavia tarkistuksia mittatiedolle:

1. Arvoalue – mittadataa verrataan suurelle asetettuihin raja-arvoihin.
2. Riippuvuus – suureille, jotka ovat riippuvuussuhteessa toisiinsa, suoritetaan testi, jotta epätodennäköiset tilanteet saadaan selville.
3. Trendi – tarkastellaan suureen pitkän aikavälin muutoksia, eli trendejä.

Näistä suoritetaan automaattisesti vain arvoalueen tarkistus. Riippuvuuksien ja trendien tarkastelu jää käyttäjälle, ja selviää usein vasta analyysin tuloksia tarkasteltaessa (etenkin pitkän aikavälin muutokset).

Jokaiselle suurelle on määritelty arvoalue, jonka rajoissa mittausten voidaan olettaa pysyvän. Arvoalue on pääsääntöisesti vakio, paitsi ilman lämpötilan yhteydessä – siinä arvoalue riippuu ajasta: tammikuulla on eri ylä- ja alarajat kuin heinäkuulla. Arvoalue tarkistetaan automaattisesti mittadatan tuonnin yhteydessä, sekä mahdolliset ylitykset ja alitukset korjataan myös automaattisesti. Olkoon L alaraja ja H yläraja ($L < H$) sekä x ja s edelleen data- ja statusvektorit. Arvoalueen tarkistus tapahtuu kahdessa vaiheessa päivittämällä ensin statusvektori:

$$s_i = \begin{cases} 2 & , x_i < L \\ 3 & , x_i > H \\ s_i & , \text{muulloin} \end{cases} , i = 1, 2, \dots, M . \quad (9)$$

On tärkeää huomata, että ala- tai ylärajan vertaaminen NaN-arvoon on aina epätosi, jolloin statusta ei muuteta. Kun statusvektori on päivitetty, voidaan mittadata korjata:

$$x_i = \begin{cases} L & , s_i = 2 \\ H & , s_i = 3 \\ x_i & , \text{muulloin} \end{cases} , i = 1, 2, \dots, M . \quad (10)$$

Taulukossa 4 on ICP:n määrittelemät sekä MetEnv-työkaluun toteutetut arvo-alueet tunnitaisille mittauksille. ICP määrittelee arvoalueet myös päivittäisille keskiarvoille (tai kertymälle sademäärän yhteydessä) [10][11]. MetEnviin ei päivittäisten arvojen tarkistusta ole kuitenkaan toteutettu, vaan mahdollisten epätodennäköisten päivittäisten arvojen arviointi on jätetty käyttäjälle, josta enemmän luvussa 4.

ICP ei määrittele ilmanpaineelle arvoaluetta, vaan taulukon 4 raja-arvot ovat samat kuin MetEnvin aiemmassa versiossa. Raja-arvot ilman suhteelliselle kosteudelle ja tuulen nopeudelle ovat niin ikään kopioitu tuosta samaisesta versiosta. Sen sijaan PAR:n raja-arvot olen arvioinut analyysien (päivittäinen maksimi) perusteella vuotta 2013 vanhemmasta, ja siten muilla keinoin validoidusta mittadatasta. Alaraja on oikein (PAR on Olkiluodossa varmasti 0 esimerkiksi yöllä joulukuussa), mutta yläraja voi olla turhan suuri.

Lämpötilalle on toteutettu ICP:n määrittymiä tiukemmat, kuukausittaiset raja-arvot, jotta mahdolliset epäilyttävät arvot havaitaan helpommin. ICP:n ehdottama vaihteluväli

on myös turhan laaja. Lämpötilaa mitataan useilta eri korkeuksilta, mutta raja-arvot pätevät kuitenkin kaikille mittauskorkeuksille.

Taulukko 4: Suureiden arvoalueet, suositellut sekä toteutetut. L on alaraja ja H yläraja. Viivalla on merkitty raja-arvot, joita ei [10] saati [11] määrittele. Ilman lämpötilalle on toteutettu kuukausittaiset raja-arvot.

Suure [yksikkö]	Suositus		Toteutus	
	L	H	L	H
Ilman lämpötila [°C]	-50	+50	1)	2)
Ilman suhteellinen kosteus [%]	0	100	60	100
Ilmanpaine [hPa]	-	-	950	1050
Maaperän lämpötila [°C]	-20	+30	-20	+30
Maaperän kosteus [%]	0	100	0	100
Sademäärä [mm]	0	50	0	50
Sadeaika [min]	0	60	0	60
Tuulen nopeus [m/s]	0	40	0	30
Tuulen suunta [°]	0	360	0	360
Kokonaissäteily [W/m ²]	0	1100	0	1100
PAR [μmol/s/m ²]	-	-	0	1500

kk	1	2	3	4	5	6	7	8	9	10	11	12
1)	-40	-40	-25	-20	-15	-15	-10	-15	-15	-20	-25	-40
2)	+20	+20	+25	+35	+40	+40	+40	+40	+35	+30	+25	+20

Kun mittadata on validoitu, siitä on korjattu mahdolliset arvoalueen ylittävät ja alittavat mittaukset, jotka on puolestaan kirjattu ylös statusvektoriin. Statusvektoriin on kirjattu myös ne alkiot mittadatasta, jotka puuttuvat, mutta puuttuvien mittausten muodostamia aukkoja ei ole vielä korjattu. Mittadata on validia vasta sitten, kun nämä aukot on korjattu ja statusvektori päivitetty asianmukaisesti. MetEnv ei anna käyttäjän aloittaa analyysiä ennen kuin kaikki mittadata on validia.

3.2 Korjausmenetelmät

Korjausmenetelmien tavoite on korvata puuttuva data parhaalla estimaatilla virheen minimoimiseksi [12]. ICP [10] ja WebMET (*The Meteorological Resource Center*) [12] määrittelevät seuraavat ”parhaat estimaattorit”:

1. Puuttuva mittaus korvataan edellisen periodin (tunnin) mittauksella. Menetelmä soveltuu yksittäisten, tunnin mittaisten aukkojen korjaamiseen, ja sitä voidaan käyttää kaikille suureille. Varovaisuutta tulee noudattaa korjattaessa periodia, jolloin siirrytään päivästä yöhön tai päinvastoin.
2. Interpolointi soveltuu yksittäisten, tunnin mittaisten aukkojen paikkaamiseen. Menetelmää voidaan käyttää myös pidempien aukkojen täyttöön tietyille suureille, kuten lämpötila. Varovaisuutta on syytä noudattaa siirtymävaiheissa tässäkin.
3. Profilointi. Menetelmässä aukko paikataan estimaatilla, joka perustuu mittadataan samankaltaisesta suureesta. Tällaisia ovat esimerkiksi OL-WOM1 -aseman eri korkeuksilta mitatut lämpötilat. Estimaatin virhe ei kasva suhteessa aukon leveyteen. Siksi profilointi on parempi kuin kumpikaan menetelmistä 1 tai 2, kun puuttuvan datan määrä kasvaa. [12]

Jos mikään edellisistä ei ole mahdollista, kohtuullisen hyvä estimaatti saavutetaan seuraavilla käytännöillä [12]:

- Korvaava data otetaan toisen mitta-aseman vastaavasta suureesta, jota mitataan läheltä, samalta korkeudelta ja samankaltaisesta maastosta. Tämä koskee lähinnä profilointia.
- Menetelmä 1, jos aukko on useamman tunnin mittainen.
- Interpolointi, jos aukko on useamman tunnin mittainen.

Estimaatti on huonoin, kun

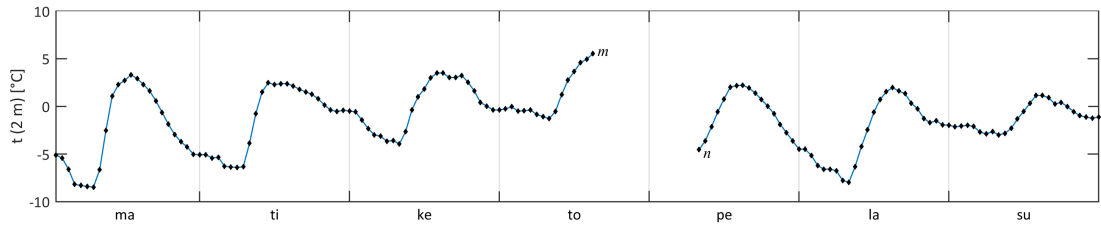
- korvaava data on lähellä olevalta mitta-asemalta, mutta erilaisesta maastosta,
- korvaavana datana käytetään jotakin analyysin tulosta, kuten esimerkiksi kuukausittaista keskiarvoa,
- korvaavana datana käytetään simulaation tulosta, tai
- korvaava data on vain jokin vakio.

Menetelmät 1, 2 ja 3 on toteutettu MetEnv-työkaluun. Ne korjaavat puuttuvien mittausten muodostamat aukot muuttamatta aitoja tai jo korjattuja mittauksia. Aina kun mittadataa korjataan jollakin menetelmistä, MetEnv päivittää statusvektoria automaattisesti menetelmälle määrätyllä statusluokalla. Näin mitta-alkioille suoritetuista operaatioista jää kommentti, sekä aidot mittaukset voidaan yhä erottaa korjatuista.

Kuvassa 2 on viikon mittainen periodi lämpötilasta (2 m) OL-WOM1 -asemalta. Mittadatassa on aukko alkaen torstain iltapäivästä ulottuen perjantain aamuun. Jatkossa heti aukkoa edeltävä indeksi on m ja heti aukkoa seuraava indeksi on n ($m < n$). Kuvassa indeksien m ja n arvot ovat 2248 ja 2265, joita vastaavat ajanhetket 4.4.2013 kello 15:00 ja 5.4.2013 kello 8:00. Indeksien perusteella voidaan laskea aukon leveys w :

$$w = n - m - 1 . \tag{11}$$

Tässä aukon leveys w on 16.



*Kuva 2: Tunneittainen lämpötila maan tasolta
1.4.2013 – 7.4.2013 väliseltä ajalta OL-WOM1 -asemalta.
Mittadatassa on aukko välillä 04.04.2013:16:00 – 05.04.2013:07:00,
yhteensä 16 tuntia.*

3.2.1 Manuaalinen korjaus

Yksinkertaisin mutta (aukon laajuudesta riippuen) työläin korjausmenetelmä on korjata mittadatan manuaalisesti. Mikäli käyttäjä korjaa jonkin data-alkion manuaalisesti, statusvektoria päivitetään korjatun alkion indeksillä arvoon 4. Muista menetelmistä poiketen käyttäjän on mahdollista korvata aito mittaus tai jo korjattu arvo manuaalisesti.

3.2.2 Vakiolla täyttö

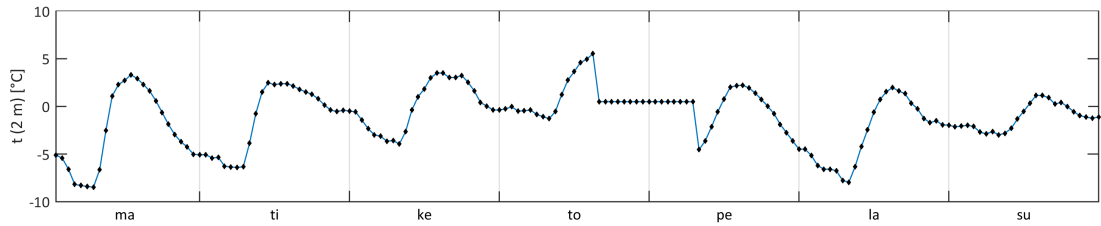
Tämä menetelmä vastaa edellä esiteltyä menetelmää 1. Olkoon c jokin reaaliarvo sekä \mathbf{x} ja \mathbf{s} määritelmien (1) ja (4) mukaiset data- ja statusvektorit. Statusvektorin perusteella menetelmä korvaa puuttuvat mittaukset vakiolla c :

$$x_i = \begin{cases} c & , s_i = 1 \\ x_i & , s_i \neq 1 \end{cases} , i = 1, 2, \dots, M. \quad (12)$$

Menetelmä päivittää statusvektoria siten, että jokainen korvattu alkio merkitään statusvektoriin arvolla 8.

Kun operaatio suoritetaan jollekin suurelle, MetEnv pyytää käyttäjää ensin syöttämään vakioarvon, jolla aukot paikataan. Kuvassa 3 on kuvan 2 mittadatan aukko korjattu tällä menetelmällä. Tässä vakiona on käytetty indeksien m ja n osoittamien mittausten keskiarvoa: $c = \frac{1}{2}(t_m + t_n) = \frac{1}{2}(5,53^\circ\text{C} - 4,54^\circ\text{C}) = 0,495^\circ\text{C}$, jossa t_i on datavektorin \mathbf{t} (lämpötila) alkio. Keskiarvo on laskettu erikseen, MetEnv ei sitä laske tämän menetelmän yhteydessä.

Työkalu on suunniteltu käsittelemään aina koko vuoden mittadatan kerrallaan, joten kaikki aukot korvataan käyttäjän antamalla arvolla. Tätä menetelmää tulee välttää tilanteissa, joissa mittadatan sisältää useita yksittäisiä aukkoja etenkin, jos ne ovat vielä ajallisesti etäällä toisistaan. Tällöin mittadatan korjaamiseen tulee harkita jotain toista menetelmää.



Kuva 3: Aukko mittadatassa on korjattu täyttämällä se vakiolla $0,5^{\circ}\text{C}$.

3.2.3 Lineaarinen interpolointi

Kaksi tason pistettä (x_a, y_a) ja (x_b, y_b) määrittävät aina yksikäsitteisen suoran, joka kulkee kummankin pisteen kautta. Näiden kahden pisteen erotusosamäärä on aina yhtä suuri kuin minkä tahansa muun, suoralla olevan kahden pisteen erotusosamäärä. Erotusosamäärä on vakio, jota sanotaan kulmakertoimeksi k . [13]

Valitaan jokin piste (x, y) suoralta siten, että $x \in (x_a, x_b)$, ja lasketaan erotusosamäärä pisteen (x_a, y_a) kanssa sekä merkitään se yhtä suureksi kuin pisteiden (x_a, y_a) ja (x_b, y_b) erotusosamäärä:

$$\frac{y - y_a}{x - x_a} = \frac{y_b - y_a}{x_b - x_a} = k, \quad x_a \neq x_b. \quad (13)$$

Yhtälöstä (13) voidaan nyt ratkaista tuntematon y :

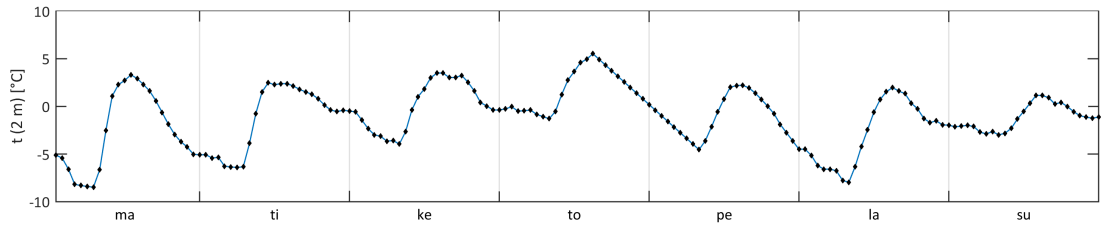
$$y = y_a + k(x - x_a). \quad (14)$$

Yhtälöllä (14) voidaan laskea y :n arvo missä tahansa kohdassa $x \in (x_a, x_b)$. Olkoon \mathbf{x} ja \mathbf{s} edelleen data- ja vastaava statusvektori. Muokataan yhtälö (14) diskreettiin muotoon käyttäen indeksejä m ja n , jotka aiemmin määriteltiin:

$$x_i = x_m + k(i - m) = x_m + \frac{x_n - x_m}{n - m}(i - m), \quad m < i < n. \quad (15)$$

Menetelmä päivittää statusvektoria siten, että jokainen interpoloitu alkio merkitään statusvektoriin arvolla 7.

Myös interpolointi suoritetaan koko mittadatalle kerrallaan. Menetelmä vaatii alkua ja loppupisteet, jolloin välittömästi vuoden alusta tai lopusta puuttuvia mittauksia joutuu ensin korjaamaan manuaalisesti. Kuvassa 4 on kuvan 2 mittadata korjattu interpoloimalla.



Kuva 4: Mittadata on korjattu interpoloimalla.

3.2.4 Profilointi

Profilointi on edellä selitetyjä menetelmiä monimutkaisempi ja vaatii referenssidataa, jota verrataan korjattavaan dataan. Referenssidatalla tarkoitetaan jotakin toista suuretta, jonka voi olettaa muuttuvan samalla tavalla kuin korjattava data. Tällaisia suureita ovat ainakin OL-WOM1 -asemalta mitatut lämpötilat maan tasolta sekä 20, 60 ja 100 metrin korkeuksilta.

Toteuttamani profilointialgoritmi käsittää seuraavat vaiheet:

1. Tarkistetaan, että referenssidataa voidaan käyttää.
2. Etsitään korjattavan mittadatan statusvektorista kaikki aukkopaidat.
3. Lasketaan keskimääräinen erotus korjattavan ja referenssidatan välillä jokaiselle aukolle erikseen.
4. Korjataan mittadata referenssidatan ja keskimääräisen erotuksen avulla.

Määritelmien (1) ja (4) mukaan, olkoon jatkossa \mathbf{x} korjattavana oleva mittadata, jonka statusvektori on \mathbf{s} , ja \mathbf{x}^{ref} referenssidata, jonka statusvektori on \mathbf{s}^{ref} . Referenssidataa voidaan käyttää profilointiin, jos sillä ja mittadatalle ei ole päällekkäisiä puuttuvia mittauksia. Tarkistus tehdään statusvektorien perusteella:

$$\forall i: \neg(s_i = 1 \wedge s_i^{\text{ref}} = 1) \equiv s_i \neq 1 \vee s_i^{\text{ref}} \neq 1, \quad i = 1, 2, \dots, M. \quad (16)$$

Jos lause (16) on tosi, algoritmi siirtyy etsimään aukkoja statusvektorin \mathbf{s} perusteella.

Toinen vaihe tuottaa matriisin $\mathbf{G}^{k \times 2}$, missä k on aukkojen lukumäärä, ja jonka sarakkeessa $\mathbf{g}_{*,1}$ on indeksit, joista aukot alkavat, ja sarakkeessa $\mathbf{g}_{*,2}$ vuorostaan indeksit, joihin aukot päättyvät. Toisin sanoen indeksipari $(g_{i,1}, g_{i,2})$ kertoo aukon i sijainnin datavektorissa \mathbf{x} . Matriisin \mathbf{G} alkioille pätee myös:

$$g_{1,1} < g_{1,2} < g_{2,1} < g_{2,2} < \dots < g_{i,1} < g_{i,2} < \dots < g_{k,1} < g_{k,2}, \quad (17)$$

jolloin datassa on oikeasti puuttuvia mittauksia, jos $k > 0$. Kun matriisi on tiedossa, algoritmi laskee keskimääräiset erotukset jokaista aukkoa varten.

Keskimääräinen erotus

Keskimääräisen erotuksen laskeminen on sinänsä yksinkertaista. Olkoon \mathbf{u} ja \mathbf{v} reaaliavaruuden vektoreita, joiden kummankin pituus on L . Keskimääräinen erotus \bar{d} on siten:

$$\bar{d} = \frac{1}{L} \sum_{i=1}^L (u_i - v_i). \quad (18)$$

Keskimääräinen erotus kertoo keskimääräisen muutoksen suuruuden lisäksi myös muutoksen suunnan.

Kuten edellä totesin, keskimääräinen erotus lasketaan jokaista aukkoa varten. Olen toteuttanut algoritmin siten, että erotuksen laskentaan huomioidaan vain aukon leveyden verran *aitoja* mittauksia ja mahdollisimman läheltä aukkoa. Näin ollen vain data, joka on ajallisesti läheltä aukkoa, huomioidaan, ja ennen kaikkea ajallisesti kaukana oleva data pyritään sivuuttamaan. Olkoon aukon i indeksipari edelleen $(g_{i,1}, g_{i,2})$. Aukon i leveys w_i on:

$$w_i = g_{i,2} - g_{i,1} + 1. \quad (19)$$

Alkioiden tai mittausten, joista keskimääräinen erotus lasketaan, valinta voidaan tehdä neljällä eri tavalla:

1. valitaan puolet alkioista juuri ennen aukkoa ja puolet juuri aukon jälkeen
2. valitaan kaikki alkioit juuri ennen aukkoa
3. valitaan kaikki alkioit juuri aukon jälkeen
4. valitaan kaikki aidot mittaukset (oletus).

Tavat ovat prioriteetin mukaisessa järjestyksessä: 1 on korkein ja 4 matalin. Tapa valitaan, mikäli seuraava ehto on tosi:

$$\forall j \in J_{i,n}: s_j = 0 \wedge s_j^{\text{ref}} = 0. \quad (20)$$

Joukko $J_{i,n}$ on niiden indeksien joukko, jolla keskimääräinen erotus lasketaan aukolle i , kun tapa n valitaan. Tavan $n = 1$ tapauksessa $J_{i,1}$ on:

$$J_{i,1} = \{ g_{i,1} - a_i, \dots, g_{i,1} - 1, g_{i,2} + 1, \dots, g_{i,2} + b_i \}, \quad (21)$$

$$\text{missä } a_i = \text{floor}\left(\frac{w_i}{2}\right) \text{ ja } b_i = w_i - a_i.$$

Keskimääräinen erotus aukolle i lasketaan siis tavalla 1, mikäli juuri ennen ja jälkeen aukkoa on sekä korjattavassa että referenssidatassa pelkästään aitoja mittauksia aukon leveyden w_i verran. Tapojen $n = 2$ ja $n = 3$ tapauksissa indeksijoukot ovat vastaavasti:

$$J_{i,2} = \{ g_{i,1} - w_i, \dots, g_{i,1} - 1 \} \text{ ja} \quad (22)$$

$$J_{i,3} = \{ g_{i,2} + 1, \dots, g_{i,2} + w_i \}. \quad (23)$$

Kun algoritmi löytää ensimmäisen indeksijoukon, $J_{i,1}$, $J_{i,2}$ tai $J_{i,3}$, joka toteuttaa ehdon (20), voidaan keskimääräinen erotus \bar{d}_i laskea aukolle i muokkaamalla kaavaa (18) seuraavasti:

$$\bar{d}_i = \frac{1}{w_i} \sum_{j \in J_{i,n}} (x_j - x_j^{\text{ref}}). \quad (24)$$

Jos ehto (20) ei täyty yhdellekään indeksijoukolle, käytetään aukolle i oletusarvoa:

$$\bar{d}_i = \bar{d}^{\text{def}}, \quad (25)$$

joka lasketaan kaikista aidoista mittauksista:

$$\bar{d}^{\text{def}} = \frac{\sum_{j=1}^M (x_j - x_j^{\text{ref}}) v(s_j; V) v(s_j^{\text{ref}}; V)}{\sum_{j=1}^M v(s_j; V) v(s_j^{\text{ref}}; V)}, \quad V = \{0\}, \quad (26)$$

missä $v(s; V)$ on määritelmän (6) funktio sekä M on data- ja statusvektorien pituus.

Tämän vaiheen tuloksena on vektori $\bar{\mathbf{d}}$, johon on laskettu k keskimääräistä erotusta. Vektori on linjassa vaiheessa 2 muodostetun matriisin \mathbf{G} kanssa siten, että jokaista matriisin indeksiparia $(g_{i,1}, g_{i,2})$, eli aukon i indeksejä, vastaa keskimääräinen erotus \bar{d}_i .

Aukon korjaaminen

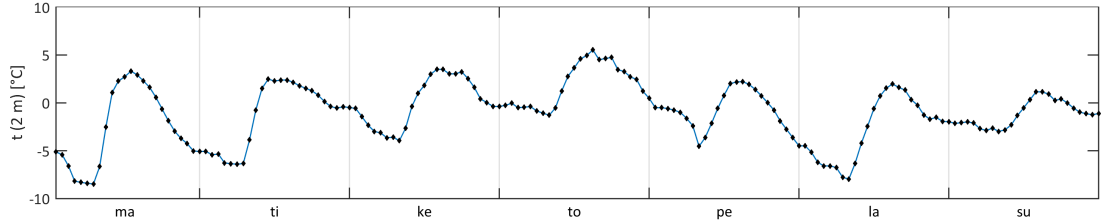
Vaiheiden 2 ja 3 jälkeen tiedossa on kaikkien aukkojen paikat (matriisi \mathbf{G}) sekä keskimääräiset erotukset (vektori $\bar{\mathbf{d}}$), joilla ne paikataan. Korjaus tapahtuu referenssidataa hyödyntäen:

$$x_j = x_j^{\text{ref}} + \bar{d}_i, \quad j = g_{i,1}, \dots, g_{i,2} \text{ ja } i = 1, \dots, k, \quad (27)$$

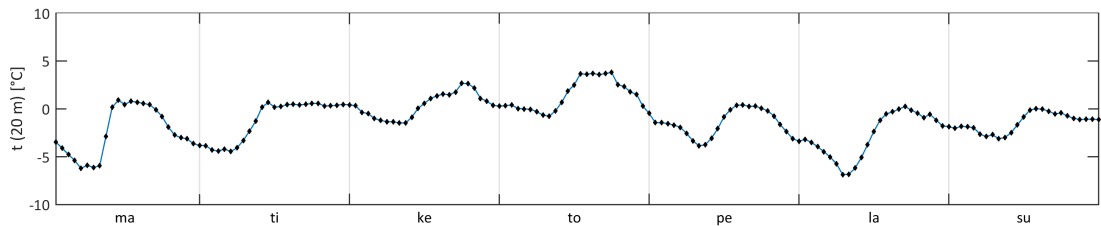
missä k on matriisin \mathbf{G} rivien, eli aukkojen lukumäärä. Kuten aiempien menetelmien tapauksissa, statusvektoria päivitetään siten, että jokainen profiloitu alkio merkitään siihen arvolla 9.

Kuvassa 5 on profiloimalla korjattu mittadata (kuva 2), joka oli siis maan tasolta mitattu lämpötila OL-WOM1 -asemalta huhtikuun alusta vuodelta 2013. Profilointiin käytetty referenssidata samalta jaksolta näkyy heti kuvassa 6. Mittatadassa on vain yksi aukko, jolle profiloointialgoritmi laski keskimääräiseksi erotukseksi $\bar{d}_1 = 0,94^\circ\text{C}$ käyttäen

tapaa 1, eli vertailu- ja mittadata valitaan ennen ja jälkeen aukon. Vertailudata on täydellistä, eli siinä ei ole yhtään aukkoa ja kaikki alkiot ovat arvoalueen sisällä.



Kuva 5: Mittadata on korjattu profiloimalla.



Kuva 6: Referenssidata: tunneittainen lämpötila 20 metrin korkeudelta 1.4.2013 – 7.4.2013 väliseltä ajalta OL-WOM1 -asemalta.

3.3 Korjausmenetelmien arviointi

Tässä luvussa arvioidaan korjausmenetelmiä. Arviointi tapahtuu tekemällä mittadataan, joka on täydellistä, aukkoja, korjaamalla aukot, ja laskemalla alkuperäisen datan ja korjatun datan välinen virhe. Paras menetelmä on luonnollisesti se, joka on lähimpänä alkuperäistä dataa, eli joka minimoi virheen.

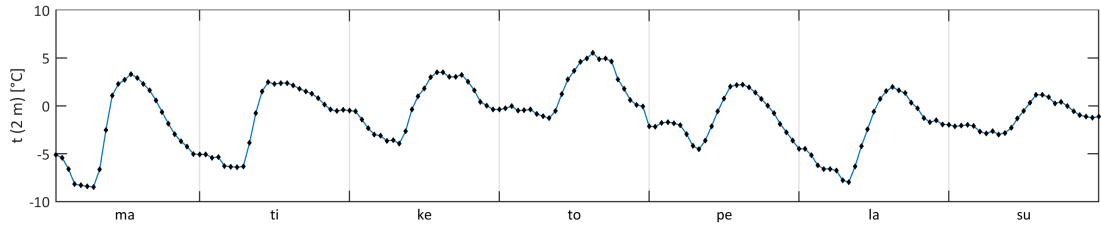
Virhe voidaan laskea, kun tiedetään alkuperäinen sekä estimoitu, eli korjattu data. Keskineliövirhe MSE (*Mean Squared Error*) on yksi tapa laskea virhe. Olkoon alkuperäinen mittadata \mathbf{x}^{org} ja korjattu \mathbf{x}^{est} , ja olkoon indeksipari $(g_{i,1}, g_{i,2})$ yhä aukon i indeksit sekä w_i aukon leveys. Aukon i korjaamisessa tehty keskineliövirhe MSE_i on näin ollen:

$$MSE_i = \frac{1}{w_i} \sum_{j=G(i,1)}^{G(i,2)} (x_j^{\text{org}} - x_j^{\text{est}})^2, \quad (28)$$

missä $\mathbf{G}(i, 1) = g_{i,1}$ ja $\mathbf{G}(i, 2) = g_{i,2}$. Keskineliövirhe eroaa keskimääräisestä erotuksesta (18) neliöön korottamisen osalta, jonka seurauksena arvo kertoo vain virheen suuruuden, ei suuntaa.

Edellisessä kappaleessa korjattavana ollut mittadata (lämpötila maan tasolta mitattuna OL-WOM1 -asemalta) on itse asiassa täydellinen koko vuoden 2013 ajalta. Dataan on tehty tarkoituksella aukko huhtikuun ensimmäiselle viikolle, kuten kuvassa 2

näky, eri menetelmien tulosten havainnollistamiseksi. Alkuperäinen mittadata kyseiseltä jaksolta on kuvassa 7.



*Kuva 7: Tunneittainen lämpötila maan tasolta
1.4.2013 – 7.4.2013 väliseltä ajalta OL-WOM1 -asemalta.*

Eri menetelmien aiheuttamat virheet on laskettu taulukossa 5 käyttäen kaavaa (28). Korjaus on suoritettu kuten kappaleissa 3.2.2, 3.2.3 ja 3.2.4 on kuvattu.

Taulukko 5: Eri korjausmenetelmien aiheuttamat virheet.

	Profilointi	Interpolointi	Vakioarvo
<i>MSE</i>	2,1	1,0	8,3

Ei ole kovinkaan yllättävää, että vakioilla täyttö osoittautuu huonoimmaksi menetelmäksi. Onhan alkuperäisessä datassa selvästi laskeva trendi aukon kohdalla, jolloin vakioarvo, vieläpä useamman tunnin ajalta, aiheuttaa suurta virhettä. Interpolointi osoittautuu selvästi parhaimmaksi menetelmäksi, joka on siinä mielessä yllättävää, että profilointi ottaa kuitenkin huomioon yksityiskohtaisia muutoksia, joita interpoloidussa datassa ei näy. Silmämääräisesti tämän havaitsee vertaamalla kuvia 5 ja 4 kuvaan 7.

Taulukko 5 ei kuitenkaan kerro vielä paljoa menetelmien hyvydestä, koska puuttuvat mittaukset eivät valikoidu satunnaisiin kohtiin toisin kuin oikeat virheet, vaan ne olivat tarkoituksella valittuja. Paremman kuvan saamiseksi suoritetaan kaksi laajempaa testiä, kun mittadatassa on useampi aukko ja niiden sijainnit satunnaisia (tietyin rajoituksin). Menetelmien hyvyttä arivoidaan testeissä keskimääräisellä keskineliövirheellä \overline{MSE} , joka on siis keskineliövirheiden keskiarvo:

$$\overline{MSE} = \frac{1}{k} \sum_{i=1}^k MSE_i, \quad (29)$$

missä k on aukkojen lukumäärä ja MSE_i määritelmän (28) mukainen keskineliövirhe.

3.3.1 Testit

Molemmassa testissä käytetään edelleen samaa mittadataa, eli maan tasolta mitattua lämpötilaa vuodelta 2013 OL-WOM1 -asemalta. Profilointiin käytetty referenssidata on niin ikään samalta asemalta ja vuodelta 2013 mitattu lämpötila 20 metrin korkeudesta. Vakio, jolla aukko täytetään, on aukon viereisten mittausten keskiarvo, joten menetelmä eroaa hieman kappaleessa 3.2.2 esitetystä.

Testi 1 suoritetaan neljälle kuukaudelle, jotka on valittu eri vuodenaajoilta. Jokaisesta päivästä kyseisiltä kuukausilta on poistettu 6 peräkkäistä mittausta siten, että aukko alkaa aina kello 15:00 ja päättyy kello 21:00. Jokaisen aukon leveys on siis 7. Testin 1 tulokset on taulukossa 6.

Testi 2a suoritetaan vuoden 2013 mittadatalle siten, että jokaiselta kuukaudelta valitaan satunnainen vuorokausi, johon tehdään aukko välille klo. 15:00 – 21:00. Mittadata sisältää siten 12 aukkoja, joiden leveys on 7 alkioita. Satunnaiset päivät ovat 9.1, 2.2, 3.3, 25.4, 21.5, 10.6, 29.7, 2.8, 14.9, 12.10, 23.11 ja 24.12. Testi 2b tehdään vertailun vuoksi myös siten, että edellä valituilta vuorokausilta poistetaan kaikki mittaukset. Testin 2 tulokset ovat myös taulukossa 6.

Testissä 1 lämpötilan pitäisi pääsääntöisesti laskea valituilla tunneilla vuodenaajasta riippumatta, jolloin myös interpoloimalla saavutetaan hyvä tulos. Kuten taulukosta 6 havaitaan, lineaarinen interpolaatio kykenee korjaamaan aukkoja keskimäärin hyvin, joskus jopa profilointia paremmin (maaliskuu). Interpolointi ei tuota merkittävästi heikompia tuloksia, vaikka vuorokaudet valittaisiin satunnaisesti, kuten testin 2a tulokset osoittavat. Sen sijaan kun data puuttuu koko vuorokauden ajalta, kuten testissä 2b, ainoastaan profilointi kykenee korjaamaan aukot hyvin, ja kuten arvata saattaa, interpolointi on jo huomattavasti heikompi menetelmä. Vakioarvolla täyttö menestyi kaikissa testitapauksissa odotetusti heikoiten.

Taulukko 6: Testien tulokset. \overline{MSE} -arvot on listattu menetelmittäin.

	Profilointi	Interpolointi	Vakioarvo
tammikuu	0,50	1,03	1,45
maaliskuu	1,33	0,82	3,52
kesäkuu	1,07	1,53	2,39
syyskuu	0,91	1,11	2,17
testi 2a	1,14	1,41	2,54
testi 2b	1,16	6,39	6,73

3.3.2 Yhteenveto korjausmenetelmistä

Edellisessä kappaleessa suoritettujen testien perusteella puuttuva mittadata kannattaa korjata profiloimalla, jos se on mahdollista (eli jos vertailudataa samalta jaksolta ja asemalta on olemassa). Vaikka testit ovatkin lähinnä suuntaa antavia, profiloimalla saavutetaan pääsääntöisesti paras tulos, eli pienin \overline{MSE} -arvo. Näin siitäkin huolimatta, vaikka testit 1 ja 2a olivatkin viritettyjä suosimaan lineaarista interpolointia. Jaksathan, jotka datasta poistettiin, olivat valittu siten, että lämpötila laskisi karkeasti ottaen monotonisesti jakson aikana. Profiloinnin etu muihin menetelmiin verrattuna käy ilmiselväksi heti, kun peräkkäisen puuttuvan mittadatan määrä kasvaa, kuten testi 2b osoittaa.

MetEnv sallii profiloinnin käytön vain saman aseman suureille. Ei ole esimerkiksi mahdollista korjata OL-WOM3 -mittadataa OL-WOM4 -aseman datalla, vaikka kyseiset asemat sijaitsevatkin lähellä toisiaan (ks. kuva 1). Profilointia voi ja kannattaa käyttää seuraavien suureiden korjaamiseen:

- Eri korkeuksilta mitatut lämpötilat OL-WOM1 -asemalta.
- Eri korkeuksilta mitatut tuulen suunnat ja nopeudet OL-WOM1 -asemalta.
- Eri korkeuksilta mitatut lämpötilat OL-WOM2 -asemalta huomioiden, että lämpötilat 2:n ja 9:n metrin korkeuksilta on mitattu latvuston sisältä toisin kuin 24:n metrin korkeudelta, joka on latvuston päältä ja voi täten aiheuttaa suurempaa virhettä.
- Eri syvyyksiltä (sekä osasta syvyyksiä useammalta anturilta) mitattavat maaperän lämpötilat asemilla 2 – 5.
- Useammalta anturilta mitattavat maaperän kosteudet asemilla 2 – 5.

Lineaarinen interpolointi soveltuu parhaiten yksittäisten tai lyhyiden aukkojen korjaamiseen etenkin, jos aukko sijaitsee jaksolla, jolloin suureen voi olettaa joko kasvavan tai vähenevän kutakuinkin monotonisesti, kuten lämpötila aamuisin ja iltaisin. Jos voidaan olettaa, että aukossa sijaitsee paikallinen minimi tai maksimi (esimerkiksi päivän kuumin tai kylmin tunti osuu aukkoon), lineaarisen interpoloinnin aiheuttamaa virhettä voidaan pienentää korjaamalla oletettu ääriarvotunti manuaalisesti ennen interpoloinnin suorittamista. Tämä tietysti vaatii tietoa alueella vallinneesta säästä kyseiseltä ajanjaksolta.

Vakiolla täyttö menestyi testeissä huonoiten. Menetelmä onkin kehitetty tilanteeseen, jossa puuttuvan peräkkäisen datan määrä on suuri, vertailudataa ei ole olemassa eikä ole mitään syytä olettaa, että suure kasvaisi tai vähenisi korjattavalla jaksolla monotonisesti. Toisin sanoen, vakiolla täyttö on viimeinen keino mittadatan korjaamiseksi.

Sademäärä sekä ilman suhteellinen kosteus OL-WOM1 -asemalta sisältävät toisinaan laajojakin aukkoja, koska anturi tai mittalaite on jäänyt ja siten epäkunnossa. Vuonna 2013 sademäärästä puuttui 992 peräkkäistä mittausta melkein heti vuoden alusta ja ilman kosteudesta jopa 1186 mittausta heti vuoden alusta. Profilointia ei voi käyttää

kumpaankaan, koska vertailudataa ei ole. On myös ilmiselvää, että interpolointi aiheuttaisi massiivista virhettä varsinkin, jos näin pitkän aukon viereiset alkiot sattuisivat olemaan nollaa suurempia. Tällöin sademäärä kyseiseltä jaksolta voisi kasvaa ennätyslukemaan. Vakiolla täyttökään tuskin minimoisi virhettä yhtään interpolointia paremmin ilman kosteuden tapauksessa, vaikka se on interpoloinnille ainoa vaihtoehto.

Sen sijaan sademäärän tapauksessa aukon täyttäminen vakiolla kuten esimerkiksi nollalla voi osoittautua jopa hyväksi ratkaisuksi, koska käyttäjän on mahdollista verrata virheellistä jaksoa muiden asemien mittadataan sekä sadeaikaan, jota OL-WOM1 -asemalta mitataan. Voi olla, että muut asemat eivät myöskään ole raportoinut yhtään sadetta, sadeaika näyttää nollaa, ja jopa ilmanpaineikin viittaa korkeapaineeseen, jolloin aukon täyttäminen nollalla olisi perusteltua.

MetEnv ei salli usean eri menetelmän yhdistelmiä mittadatan korjaamiseksi, vaan kun jokin menetelmä suoritetaan mittadatalle, kaikki aukot korjataan samalla menetelmällä koko vuoden ajalta. Näin menetellään, koska profilointi minimoi virhettä keskimäärin parhaiten, joten ei ole syytä vaihtaa menetelmää kesken vuoden, ja koska halusin pitää toteutuksen yksinkertaisena. Mikäli työkalua halutaan jatkossa kuitenkin kehittää edelleen, yhdistelmien salliminen voi olla hyödyllistä.

4 ANALYYSI

Mitta-asemien tuottama ja MetEnvin käyttäjän sekä korjaama että validoima mittadata on usein hyvin pitkä, yhtäjaksoinen aikasarja mittauksia, joista yksistään on vielä hyvin hankalaa muodostaa kuvaa Olkiluodon saarella kulloinkin vallinneista säätiloista. Siksi mittadataa on sopivasti ensin analysoitava.

Analyysissä suureista lasketaan muun muassa erilaisia tilastollisia tunnuslukuja, kuten vaikkapa keski- ja ääriarvoja. Näiden tunnuslukujen perusteella voidaan laskea monimutkaisempiakin lukuja, kuten vaikkapa vallitseva tuulen suunta. Analyysiä toteutetaan vuotuisille, kuukausittaisille ja päivittäisille jaksoille, sekä myös pitkälle aikavälille kattaen kaikki mitatut vuodet. Analyysi on automaattinen prosessi, joka ei pääasiassa vaadi MetEnvin käyttäjältä erillistä panosta.

4.1 Laskentakanta ja analysoitavat jaksot

Laskentakanta on tietovarasto, johon validi mittadata tallennetaan. Kanta on paikallinen, eli se sijaitsee tietokoneella, jossa MetEnviä suoritetaan. Koska kannan data on validia, se on yhtäjaksoista, josta kaikki mahdolliset aukot on korjattu, ja joka on siten valmis analysoitavaksi.

Mittadataa on kerrytetty suureille useiden vuosien ajalta yhtäjaksoisesti aina tunnin välein. Taulukossa 7 on kaikilta mitta-asemilta kertyneen datan määrä vuoden 2013 loppuun mennessä. Esimerkiksi OL-WOM1 -asemalta mittadataa on 184080 alkiota jokaisesta mitatusta suureesta, joista kaikki eivät tosin ole aitoja mittauksia. Mittadataa on itse asiassa vielä tätäkin enemmän, koska taulukkoon, kuten myös analyysiin, huomioidaan vain ne vuodet, joilta dataa on koko vuoden ajalta. OL-WOM1 -asemalta olisi dataa saatavilla jo 14.1.1992 kello 11:00 lähtien, mutta kaikki vuoden 1992 mittadata kuitenkin sivuutetaan analyyseissä. Tämä pätee myös muihin asemiin, joilta kaikilta olisi dataa saatavilla jo aikaisemmin kuin mitä taulukossa 7 on lueteltu.

Taulukko 7: Mittadatan kertymät asemittain vuoden 2013 loppuun mennessä.

OL-WOM _n	1	2	3	4	5
1. vuosi (y_1)	1993	2005	2006	2008	2010
Kertymä	184080	78888	70128	52608	35064

Kertymä on siis tuntien lukumäärä, joka on kertynyt 1. vuoden alusta vuoden 2013 loppuun. Kun mittadatan alku- ja loppupisteiden aikaleimat, eli päivämäärä ja tunti tiedetään, sekä tiedetään, että dataa on aina tunnin välein, voidaan mittadata pilkkoa ajallisesti lyhyempiin osiin. Mittadatasta analysoidaankin vuosien, kuukausien ja vuorokausien pituisia jaksoja.

Kaikki laskentakannassa olevat suureet analysoidaan automaattisesti. Analyysi tapahtuu kahdessa vaiheessa, joista ensimmäisessä vaiheessa mittadatalle suoritetaan perusanalyysi. Siinä datasta lasketaan jakson keskiarvoja, ääriarvoja ja toisinaan kertymiä, sekä mittadatan eheys. Toisen vaiheen analyysissä perusanalyysin tuloksista lasketaan joillekin suureille tiettyjen ehtojen täyttämien päivien lukumäärää sekä pitkän aikavälin analyysi.

4.1.1 Analysoitavat jaksot

Laskentakannassa olevan suureen mittadata on määritelmän (1) vektori, jonka pituus M on taulukon 7 mukainen. Jokaista vektorin indeksiä vastaa tietty aikaleima, joka saadaan selville, kun ensimmäiseen indeksiin, eli indeksiin 1 sidottu aikaleima tiedetään. Kun tiedetään lisäksi, että vuodessa on 8760 tuntia ja karkausvuodessa 8784 tuntia, voidaan laskea ne indeksit, joista vuodet alkavat ja joihin ne päättyvät. Ensimmäiseen indeksiin sidottu aikaleima on tallennettu laskentakantaan aina jokaisen suureen yhteeseen.

Kokonaisia vuosia on kertynyt esimerkiksi mitta-asemalta OL-WOM5 neljä vuoden 2013 loppuun mennessä, joista 2012 oli karkausvuosi. Näin ollen vastaava vuosien indeksimatriisi \mathbf{Y} kaikille mitta-aseman 5 suureille on:

$$\mathbf{Y} = \begin{pmatrix} 1 & 8760 \\ 8761 & 17520 \\ 17521 & 26304 \\ 26305 & 35064 \end{pmatrix}, \quad (30)$$

missä sarake $\mathbf{y}_{*,1}$ sisältää indeksit, joista vuodet alkavat, ja $\mathbf{y}_{*,2}$ indeksit, joihin ne päättyvät. Varsinaiset vuosiluvut y_i saadaan yksinkertaisella laskulla:

$$y_i = y_1 - 1 + i \iff i = y_i - y_1 + 1, \quad (31)$$

missä y_i on matriisin \mathbf{Y} riviä i vastaava vuosiluku ja y_1 ensimmäinen vuosiluku (ks. taulukko 7), joka aseman OL-WOM5 tapauksessa on siis 2010.

Kuukausien indeksimatriisi ei riipu mitta-asehasta, vaan vuodesta, jolle matriisi lasketaan. Jos vuosi ei ole karkausvuosi ja helmikuussa on 28 päivää, kuukausien indeksimatriisi \mathbf{M} on:

$$\mathbf{M}^T = \begin{pmatrix} 1 & 745 & 1417 & 2161 & 2881 & 3625 & 4345 & 5089 & 5833 & 6553 & 7297 & 8017 \\ 744 & 1416 & 2160 & 2880 & 3624 & 4344 & 5088 & 5832 & 6552 & 7296 & 8016 & 8760 \end{pmatrix}, \quad (32)$$

missä sarake $\mathbf{m}_{*,1}$ sisältää indeksit, joista kuukaudet alkavat, ja $\mathbf{m}_{*,2}$ indeksit, joihin ne päättyvät (kun matriisi ei ole transponoidussa muodossa). Jos vuosi on karkausvuosi ja helmikuussa on 29 päivää, kuukausien indeksimatriisi \mathbf{M}_{ly} on vastaavasti:

$$\mathbf{M}_{ly}^T = \begin{pmatrix} 1 & 745 & 1441 & 2185 & 2905 & 3649 & 4369 & 5113 & 5857 & 6577 & 7321 & 8041 \\ 744 & 1440 & 2184 & 2904 & 3648 & 4368 & 5112 & 5856 & 6576 & 7320 & 8040 & 8784 \end{pmatrix}, \quad (33)$$

missä ly merkitsee karkausvuotta (*leap year*).

Myös vuorokausien indeksimatriisi riippuu vain siitä, onko kyseinen vuosi karkausvuosi vai ei. Jos vuosi ei ole karkausvuosi, vuorokausien indeksimatriisi \mathbf{D} on:

$$\mathbf{D} = \begin{pmatrix} 1 & 24 \\ 25 & 48 \\ \vdots & \vdots \\ 8737 & 8760 \end{pmatrix}, \quad (34)$$

missä sarake $\mathbf{d}_{*,1}$ sisältää indeksit, joista vuorokaudet alkavat, ja $\mathbf{d}_{*,2}$ indeksit, joihin ne päättyvät. Karkausvuoden indeksimatriisi saadaan yksinkertaisesti lisäämällä matriisiin (34) yksi vuorokausi, joten indeksimatriisi \mathbf{D}_{ly} on:

$$\mathbf{D}_{ly} = \begin{pmatrix} \mathbf{d}_{*,1} & \mathbf{d}_{*,2} \\ 8761 & 8784 \end{pmatrix}, \quad (35)$$

missä ly merkitsee edelleen karkausvuotta.

Kuukausien indeksimatriisit (32) ja (33) sekä vuorokausien indeksimatriisit (34) ja (35) ovat vakioita ja niillä voidaan indeksoida minkä tahansa vuoden mittadata kuukausien tai vuorokausien pituisiin osiin, kun tiedetään onko kyseessä karkausvuosi vai ei. Kaikilla indeksimatriiseilla, mukaanlukien vuosien matriisit, voidaan tietysti indeksoida myös mittadataa vastaavat statusvektorit.

Jatkossa indeksimatriisit (30), (32), (33), (34) ja (35) korvataan analysoitavien jaksojen yleisellä indeksimatriisilla $\mathbf{P}^{k \times 2}$, jonka sarake $\mathbf{p}_{*,1}$ sisältää jaksojen alkuindeksit ja $\mathbf{p}_{*,2}$ vastaavasti loppuindeksit, ja jossa k on analysoitavien jaksojen lukumäärä. Näin ollen kuukausittaisen analyysin yhteydessä k olisi 12 ja yleisen matriisin sisältö vastaisi joko matriisia (32) tai (33) riippuen vuodesta. Päivittäisen analyysin yhteydessä – riippuen edelleen vuodesta – k olisi joko 365 tai 366 ja yleisen matriisin sisältö vastaisi joko matriisia (34) tai (35). Sen sijaan vuotuisen analyysin yhteydessä yleisen indeksi-

matriisiin sisältö riippuisi mitta-asemasta ja se muodostettaisiin samoin periaattein kuin matriisi (30). Myös merkitätapaa $\mathbf{P}(i, j) = p_{i,j}$ käytetään tarvittaessa alaindeksien sijaan selvyyden vuoksi.

Vuotuisten, kuukausittaisten ja päivittäisten jaksojen lisäksi mittadatasta lasketaan vielä pitkän aikavälin analyysi. Se suoritetaan muun analyysin yhteydessä ja perustuu vuotuisen analyysin tuloksiin, eikä sitä varten tarvita omaa indeksimatriisia. Pitkän aikavälin analyysi käsitellään luvussa 4.2.4.

4.2 Analyysi

Analyysi suoritetaan kahdessa vaiheessa aina yhdelle suurelle kerrallaan. Analyysi ei siis riipu kuin yhden suureen mittadatasta (poikkeuksena vallitseva tuulen suunta, joka käsitellään omassa luvussaan 4.4). Suurelle suoritetaan ensiksi perusanalyysi, jonka pohjalta lasketaan toisen vaiheen analyysit, jos suurelle on sellaisia määritelty. Pitkän aikavälin analyysi lasketaan viimeisenä. Analyysin suorittaa MetEnviin toteuttamani algoritmi, jonka toiminta voidaan kuvata karkealla tasolla seuraavilla kuudella kohdalla:

1. Lasketaan vuosien indeksimatriisi (30), joka riippuu mitta-asemasta, ja käydään läpi kaikki vuodet.
2. Poimitaan mittadata- ja statusvektoreista vuotta y_i vastaavat osavektorit. Olkoon ne jatkossa \mathbf{x} ja \mathbf{s} vastaavasti.
3. Päivittäinen analyysi: lasketaan vuodelle y_i vuorokausien indeksimatriisi (34) tai (35), joilla vektoreille \mathbf{x} ja \mathbf{s} lasketaan perusanalyysi. Toisen vaiheen analyysijä ei lasketa päivittäisille jaksoille ollenkaan.
4. Vuotuinen analyysi: lasketaan vektoreille \mathbf{x} ja \mathbf{s} ensin perusanalyysi, jonka tulosten pohjalta lasketaan toisen vaiheen analyysit.
5. Kuukausittainen analyysi: lasketaan vuodelle y_i kuukausien indeksimatriisi (32) tai (33), joilla vektoreille \mathbf{x} ja \mathbf{s} lasketaan ensin perusanalyysi, jonka tulosten pohjalta lasketaan toisen vaiheen analyysit.
6. Lopuksi, kun kaikki vuodet on käyty läpi, lasketaan pitkän aikavälin analyysi vuotuisen analyysin pohjalta.

Kuten listasta havaitaan, toisen vaiheen analyysijä lasketaan ainoastaan vuotuisille ja kuukausittaisille jaksoille. Pitkän aikavälin analyysi lasketaan vuotuisen perusanalyysin ja toisen vaiheen analyysien tuloksista.

4.2.1 Perusanalyysi

Suurelle lasketaan kaikilta jaksoilta ainakin osa seuraavista tunnusluvuista: keskiarvo, ääriarvot, kertymä (tai summa) ja mittadatan eheys, joka käsitellään erikseen omassa luvussa 4.2.2. Analyysin tuloksena saadaan siis suureen vuotuiset, kuukausittaiset ja päivittäiset keskiarvot, ääriarvot ja kertymät, sekä tieto mittadatan eheydestä.

Olkoon suuretta x vastaava validi mittadatavektori \mathbf{x} ja statusvektori \mathbf{s} . Keskiarvovektorin $\bar{\mathbf{x}}^{\mathbf{P}}$ alkioit yleisille jaksoille \mathbf{P} lasketaan näin ollen:

$$\bar{x}_i^{\mathbf{P}} = \frac{\sum_{j=\mathbf{P}(i,1)}^{\mathbf{P}(i,2)} x_j v(s_j; V)}{\sum_{j=\mathbf{P}(i,1)}^{\mathbf{P}(i,2)} v(s_j; V)}, \quad i = 1, 2, \dots, k, \quad (36)$$

missä V on määritelmän (5) joukko ja $v(s; V)$ määritelmän (6) funktio. Keskiarvovektorin yläindeksi \mathbf{P} tarkoittaa eri asiaa kuin yleinen indeksimatriisi \mathbf{P} . Yläindeksi \mathbf{P} tulee korvatuksi analysoitavaa jaksoa vastaavalla kirjaimella Y , M tai D . Vuotuinen keskiarvovektori olisi siis $\bar{\mathbf{x}}^Y$ ja vastaava indeksimatriisi (30). Keskiarvoksi asetetaan 0 tilanteissa, joissa jaksoilta ei ole olemassa yhtään pätevää mittausta, eli kaavan (36) nimittäjä olisi 0.

Kertymä- tai summavektorin $\Sigma \mathbf{x}^{\mathbf{P}}$ alkioit yleisille jaksoille \mathbf{P} on kaavan (36) osoittaja:

$$\Sigma x_i^{\mathbf{P}} = \sum_{j=\mathbf{P}(i,1)}^{\mathbf{P}(i,2)} x_j v(s_j; V), \quad i = 1, 2, \dots, k. \quad (37)$$

Edelleen, päivittäinen kertymävektori olisi $\Sigma \mathbf{x}^D$ ja vastaava indeksimatriisi (34) tai (35) riippuen siitä, onko vuosi karkausvuosi vai ei.

Ääriarvoilla tarkoitetaan jaksojen maksimeja ja minimejä, sekä myös niitä indeksien arvoja, joissa maksimit ja minimiit ovat. Koska indeksiin on sidottu tietty aikaleima, maksimin tai minimin ajanhetki voidaan siten määrittää tunnilleen.

Maksimivektorin $\max \mathbf{x}^{\mathbf{P}}$ alkioit yleisille jaksoille \mathbf{P} lasketaan:

$$\max x_i^{\mathbf{P}} = \max_{j \in J_i} \mathbf{x}, \quad i = 1, 2, \dots, k \quad (38)$$

ja maksimeja vastaavan indeksivektorin $\max \mathbf{j}^{\mathbf{P}}$ alkioit ovat:

$$\max j_i^{\mathbf{P}} = \arg \max_{j \in J_i} \mathbf{x}, \quad i = 1, 2, \dots, k, \quad (39)$$

joissa kummassakin J_i on kaikkien mahdollisten pätevien indeksien joukko jaksolle i , joka määritetään statusvektorin \mathbf{s} perusteella:

$$J_i = \{ j \mid \forall j \in \{ p_{i,1}, \dots, p_{i,2} \} : s_j \in V \} \quad (40)$$

missä V on pätevien statusluokitusten, eli määritelmän (5) joukko.

Minimivektorin $\min \mathbf{x}^P$ alkioit yleisille jaksoille \mathbf{P} lasketaan:

$$\min x_i^P = \min_{j \in J_i} \mathbf{x} \quad , \quad i = 1, 2, \dots, k \quad (41)$$

ja minimejä vastaavan indeksivektorin $\min \mathbf{j}^P$ alkioit ovat:

$$\min j_i^P = \arg \min_{j \in J_i} \mathbf{x} \quad , \quad i = 1, 2, \dots, k, \quad (42)$$

missä J_i on määritelmän (40) mukainen indeksijoukko.

Kaikki funktioista (38), (39), (41) ja (42) palauttavat vain yhden arvon. Funktioiden (39) ja (42) tapauksissa palautetaan aina ensimmäinen indeksi, jonka kyseiset funktiot löytävät. Tilanteissa, joissa jaksoilta ei ole olemassa päteviä mittauksia, palautetaan funktioiden (38) ja (41) tapauksissa 0, ja funktioiden (39) ja (42) tapauksissa 1 (eli ensimmäinen indeksi).

Kaikkia perusanalyysejä, joita tässä luvussa on määritelty, ei lasketa jokaiselle suureelle. Taulukkoon 8 on merkitty kullekin suureelle laskettavat perusanalyysit. Esimerkiksi kertymää ei lasketa kuin sademäärälle ja -ajalle, kun taas tuulen suunnalle (tai leviämisuunnalle) ei lasketa mitään näistä perusanalyyseistä. Toisaalta joillekin suureille lasketaan turhaan minimiarvoja, jotka ilmiön luonteen perusteella eivät ole kovinkaan kiinnostavia. Tällaisia suureita ovat tuulen nopeus, kokonaissäteily ja PAR. Näin on menetelty, koska analyysien yli hyppääminen vain monimutkistaisi toteutusta saavuttamatta kuitenkaan merkittävää etua suoritusnopeudessa. Sen sijaan, kun jokin perusanalyysi lasketaan suureelle, se lasketaan aina kaikille jaksoille, eli vuosille, kuukausille ja vuorokausille.

Taulukko 8: Suureille laskettavat perusanalyysit.

Suure [yksikkö]	Keskiarvo	Ääriarvot	Kertymä
Ilman lämpötila [°C]	X	X	
Ilman suhteellinen kosteus [%]	X	X	
Ilmanpaine [hPa]	X	X	
Maaperän lämpötila [°C]	X	X	
Maaperän kosteus [%]	X	X	
Sademäärä [mm]			X
Sadeaika [min]			X
Tuulen nopeus [m/s]	X	X	
Tuulen suunta [°]			
Tuulen suunnan hajonta [°]	X	X	
Kokonaissäteily [W/m ²]	X	X	
PAR [μmol/s/m ²]	X	X	

4.2.2 Mittadatan eheys

Analyysiä ei voida pitää luotettavana, jos mittadataa, josta analyysi lasketaan, on puuttunut suhteellisen paljon ja sitä on siten jouduttu korjaamaan. Luotettavuuden arvioimiseksi jokaiselle suurelle lasketaan aina perusanalyysin yhteydessä suhdeluku, joka kertoo jakson aitojen ja odotettujen mittausten lukumäärien suhteen. Tätä suhdelukua kutsutaan mittadatan eheydeksi, joka esiintyy MetEnvin ja analyysin tulosten yhteydessä myös lyhenteellä DI (*Data Integrity*). Se ilmoitetaan toisinaan myös prosentteina suhdeluvun sijaan.

Olkoon suureen x validia mittadatavektoria vastaava statusvektori s . Näin ollen mittadatan eheydet $_{DI}x_i^P$ yleisille jaksoille \mathbf{P} on:

$$_{DI}x_i^P = \frac{\sum_{j=\mathbf{P}(i,1)}^{\mathbf{P}(i,2)} v(s_j; \{0\})}{\mathbf{P}(i,2) - \mathbf{P}(i,1) + 1}, \quad i = 1, 2, \dots, k, \quad (43)$$

missä $v(s; \{0\})$ on määritelmän (6) funktio ja k jaksojen lukumäärä. Pätevien statusluokkien joukko on tässä vakio sisältäen vain yhden statusluokan: aidon mittauksen.

Analyysi on sitä luotettavampi mitä lähempänä mittadatan eheys on yhtä, eli mitä enemmän kyseiseltä jaksolta on olemassa aitoja mittauksia, tai päinvastoin, sitä epäluotettavampi mitä lähempänä eheys on nollaa, eli mitä vähemmän kyseiseltä jaksolta on olemassa aitoja mittauksia. Jos kaikki jakson mittaukset ovat aitoja,

mittadatan eheys on 1. Eheys voi toki olla myös 0, jolloin jaksolta ei ole saatavilla yhtään aitoa mittausta.

[10] ja [11] määrittelevät kynnyksarvon mittadatan eheydelle, joka pitää vähintään saavuttaa, kun päivittäisiä tai myös tunnitaisia arvoja lasketaan, jotta niitä voidaan pitää luotettavina. Näytteistyshän tapahtuu 10 minuutin välein kuten kappaleessa 2 kuvataan, joista sitten mitta-alkion tunnittainen arvo lasketaan joko näytteiden keskiarvona, maksimina, miniminä tai summana suureesta riippuen. Koska MetEnvin käyttäjä ei voi vaikuttaa varsinaiseen näytteistykseen, vaan ainoastaan niistä laskettuihin tunnitaisiin arvoihin, keskityn tässä vain kynnyksarvoihin, jotka on asetettu päivittäisille arvoille.

Taulukossa 9 on esitetty eheyden kynnyksarvot päivittäisille arvoille suureittain. Tämä tarkoittaa, että esimerkiksi laskettaessa ilman lämpötilan päivittäisiä keskiarvoja, aitoja mittauksia tulee olla vähintään 95% vuorokaudesta, eli käytännössä 23:lta tunnilta, jotta keskiarvo voidaan hyväksyä. Kynnyksarvot riippuvat suureen odotetusta vaihtelusta: jos suurelle on ominaista suuri vaihtelu, eheyden kynnyksarvo on tiukempi. Jos vaihtelu on sen sijaan pientä, myös kynnyksarvo on väljempi [10]. Taulukosta käy ilmi, että eheydelle asetetut kynnyksarvot ovat pääsääntöisesti melko tiukat.

*Taulukko 9: Mittadatan eheyden kynnyksarvot suureittain
kuten ICP [10] ne määrittelee.*

Suure	Kynnyksarvo
Ilman lämpötila	95%
Ilman suhteellinen kosteus	95%
Maaperän lämpötila	95%
Maaperän kosteus	50%
Sademäärä	100%
Tuulen nopeus	95% (100% maksimiarvoille)
Tuulen suunta	95%
Kokonaissäteily	100% valoiselta ajalta

MetEnvin toteutus eroaa tästä periaatteellisesti. Ensinnäkin, kuten aiemmin todettu, kaikki perusanalyysit lasketaan jokaiselle suurelle, eikä tässä vaiheessa datan eheyteen kiinnitetä huomiota, etenkin kun se vasta lasketaan osana perusanalyysiä. Toiseksi, sekä kuukausittaiset että vuotuiset analyysit lasketaan suoraan tunnitaisista eikä päivittäisistä arvoista kuten [10] antaa ymmärtää. Tällöin analyysiin tulee mukaan mittauksia päiviltä, jotka eivät ylitä suurelle asetettua kynnyksarvoa, sen sijaan, että kuukausittainen tai vuotuinen analyysi laskettaisiin vain riittävän eheyden päivien perusteella. Kolmanneksi, MetEnv ei tue suurekohtaisia kynnyksarvoja.

MetEnviin on asetettu kaksi kynnyksarvoa mittadatan eheydelle, joita käytetään toisen vaiheen ja pitkän aikavälin analyysien yhteydessä. Mikäli kynnyksarvo saavutetaan tai ylitetään, kyseinen perusanalyysin tulos otetaan huomioon. Määritellään tätä varten funktio $\theta: \mathbb{R} \rightarrow \{0, 1\} \subset \mathbb{R}$, joka huomioi kynnyksarvon Θ saavuttavat ja ylittävät arvot:

$$\theta(x; \Theta) = \begin{cases} 1 & , x \geq \Theta \\ 0 & , x < \Theta \end{cases} . \quad (44)$$

Tämä funktio on hyvin samankaltainen kuin määritelmän (6) pätevät mittaukset huomioiva funktio.

Vaikka MetEnv hyödyntää mittadatan eheyttä vain näihin kahteen edellä mainittuun analyysiin, se on kuitenkin erittäin hyödyllinen luku käyttäjän arvioidessa minkä tahansa analyysin tulosta ja sitä kannattaakin käyttää hyväksi tuloksia raportoitaessa. Raporteissa on usein mainintoja siitä, mitkä jaksot on jätetty pois, koska aidon datan määrä on ollut liian pieni ja tulos siten epäluotettava. Mittadatan eheys vaikuttaa suoraan luotettavuuteen ja se voidaan näin ollen raportoida esimerkiksi tulosten yhteydessä.

4.2.3 Toisen vaiheen analyysit

Toisen vaiheen analyysit pohjautuvat päivittäisten perusanalyysien tuloksiin, jotka on laskettu kuten kappaleessa 4.2.1 kuvataan. Toisessa vaiheessa osalle suureista lasketaan jokaiselta vuodelta niiden päivien, jotka täyttävät tietyn ehdon, lukumäärät ottaen huomioon myös mittadatan eheys. Toisen vaiheen analyysi lasketaan seuraaville suureille kaikilta mahdollisilta mitta-asemilta: ilman lämpötila, sademäärä ja tuulen nopeus. Toisen vaiheen analyysit ehtoineen on lueteltu taulukossa 10 kuten ne on Posivan raportissa [6] määritelty.

Taulukko 10: Toisen vaiheen analyysit suureittain.

Suure	Päivien lukumäärä, joina
Ilman lämpötila (kaikki korkeudet) – merkitään jatkossa symbolilla t	a) maksimiarvo on yli $+25^{\circ}\text{C}$ b) maksimiarvo on alle 0°C c) minimiarvo on alle 0°C d) minimiarvo on alle -10°C
Sademäärä – merkitään jatkossa symbolilla p	a) kertymä on vähintään 0,1 mm b) kertymä on vähintään 1,0 mm c) kertymä on vähintään 10,0 mm
Tuulen nopeus (kaikki korkeudet) – merkitään jatkossa symbolilla w	a) maksimiarvo on yli 10 m/s b) maksimiarvo on yli 20 m/s

Suureita on siis kolmea tyyppiä (eri asemilta ja korkeuksilta tehdyt mittauksethan katsotaan omiksi suureiksi, vaikka itse mitattava ilmiö olisikin sama), joista ilman lämpötiloille lasketaan neljä, sademäärille kolme ja tuulen nopeuksille kaksi toisen vaiheen analyysiä.

Olkoon X_i^{cond} niiden perusanalyysivektorin $\mathbf{x}^{i,D}$ indeksien muodostama joukko vuodelta y_i , jotka täyttävät jonkin taulukon 10 ehdon (ilmaistaan yläindeksillä cond), joka on asetettu suurelle x . Esimerkiksi joukko T_i^a vuodelta y_i , johon liittyvä suure on siis ilman lämpötila tietyltä korkeudelta ja mitta-asemalta, sisältää ne indeksit, joissa vektorin $\max \mathbf{t}^{i,D}$ alkiot ovat suurempia kuin 25. Kaikki taulukon 10 suureiden ja ehtojen muodostamat joukot on lueteltu taulukossa 11. Siinä näkyvät myös ne päivittäiset perusanalyysissä lasketut vektorit vuodelta y_i , joista joukot muodostetaan. Jokaista joukkoa on yhtä monta kuin vuosia, joilta mittauksia on tehty. Merkitään tätä lukumäärää jatkossa kirjaimella N , jonka arvo riippuu mitta-asemasta taulukon 7 mukaisesti.

Jotta mittadatan eheys voidaan huomioida, määritellään pätevien indeksien joukko J_i vuodelle y_i , jonka alkiot ovat indeksejä, joissa määritelmän (43) eheysvektorin $\text{DI}\mathbf{x}^{i,D}$ alkiot vuodelta y_i ovat vähintään yhtä suuria kuin ennalta asetettu kynnsarvo. Olkoon suureeseen x liittyvä pätevien indeksien joukko J_i^x , joka on:

$$J_i^x = \left\{ j \mid \forall j : \text{DI}x_j^{i,D} \geq \Theta \right\}, \quad i = 1, 2, \dots, N, \quad (45)$$

missä Θ on edellä mainittu kynnsarvo. MetEnvissä kynnsarvoksi on asetettu 0,95 (tai 95%) ja sitä käytetään kaikille suureille, joille toisen vaiheen analyysi lasketaan.

Taulukko 11: Suureiden ja ehtojen muodostamat joukot vuodelle y_i .

Suure x	Ehdon täyttävä joukko X_i^{cond}
Ilman lämpötila t (kaikki korkeudet). Analysoitavana vektorina on päivittäiset maksimiarvot $\max \mathbf{t}^{i,D}$ ja minimiarvot $\min \mathbf{t}^{i,D}$ vuodelta y_i .	$T_i^a = \left\{ j \mid \forall j : \max t_j^{i,D} > 25 \right\}$
	$T_i^b = \left\{ j \mid \forall j : \max t_j^{i,D} < 0 \right\}$
	$T_i^c = \left\{ j \mid \forall j : \min t_j^{i,D} < 0 \right\}$
	$T_i^d = \left\{ j \mid \forall j : \min t_j^{i,D} < -10 \right\}$
Sademäärä p . Analysoitavana vektorina on päivittäiset kertymät $\Sigma \mathbf{p}^{i,D}$ vuodelta y_i .	$P_i^a = \left\{ j \mid \forall j : \Sigma p_j^{i,D} \geq 0,1 \right\}$
	$P_i^b = \left\{ j \mid \forall j : \Sigma p_j^{i,D} \geq 1 \right\}$
	$P_i^c = \left\{ j \mid \forall j : \Sigma p_j^{i,D} \geq 10 \right\}$
Tuulen nopeus w (kaikki korkeudet). Analysoitavana vektorina on päivittäiset maksimiarvot $\max \mathbf{w}^{i,D}$ vuodelta y_i .	$W_i^a = \left\{ j \mid \forall j : \max w_j^{i,D} > 10 \right\}$
	$W_i^b = \left\{ j \mid \forall j : \max w_j^{i,D} > 20 \right\}$

Kaikki taulukon 11 sekä määritelmän (45) joukot muodostetaan päivittäisten perusanalyysien pohjilta. Näin ollen joukko X_i^{cond} on joko lukujonon $\{1, 2, \dots, 365\}$ tai $\{1, 2, \dots, 366\}$ osajoukko riippuen siitä, onko vuosi y_i karkausvuosi vai ei. Sama pätee määritelmän (45) indeksijoukolle J_i^x . Varsinaiset toisen vaiheen analyysit vuodelle y_i voidaan nyt laskea näiden kahden joukon leikkauksen kardinaliteettina.

Olkoon vektori $\mathbf{c}^{x,\text{cond}}$ jokin suurelle x määritelty taulukon 10 mukainen toisen vaiheen analyysi, joka lasketaan vuosille $y_1 \dots y_N$:

$$\mathbf{c}^{x,\text{cond}} \in \mathbb{Z}^N, \text{ missä } c_i = |X_i^{\text{cond}} \cap J_i^x|. \quad (46)$$

Esimerkiksi ilman lämpötilalle vaikkapa OL-WOM5 -asemalta lasketaan toisen vaiheen analyysissä päivien, joina maksimilämpötila ylittää $+25^\circ\text{C}$, lukumäärä kaavalla (46) seuraavasti:

$$\mathbf{c}^{t,a} \in \mathbb{Z}^4, \text{ missä } c_i = |T_i^a \cap J_i^t|,$$

missä T_i^a on taulukon 11 mukainen joukko ja J_i^t suuretta vastaava määritelmän (45) mukainen indeksijoukko, molemmat vuodelta y_i . Analysoitavia vuosia kyseiseltä asemalta on 4 alkaen vuodesta 2010 (ks. taulukko 7).

Edellä on laskettu vuotuiset toisen vaiheen analyysit, mutta nämä analyysit lasketaan myös kuukausille. Tämä tapahtuu yhä samalla periaatteella, paitsi että joukkoa, jonka kardinaliteettia lasketaan, kavennetaan koskemaan vain tiettyä kuukautta koko vuoden sijaan. Tätä varten kaavaa (46) on muokattava tähän soveltuvaksi.

Olkoon matriisi $\mathbf{C}^{x,\text{cond}}$, jossa on N riviä ja 12 saraketta, jokin suurelle x määritelty taulukon 10 mukainen toisen vaiheen kuukausittainen analyysi kattaen vuodet $y_1 \dots y_N$. Matriisi on näin ollen:

$$\mathbf{C}^{x,\text{cond}} \in \mathbb{Z}^{N \times 12}, \text{ missä } c_{i,j} = |X_i^{\text{cond}} \cap J_i^x \cap M_{i,j}|. \quad (47)$$

Joukko $M_{i,j}$ on kuukauden j indeksijoukko vuodelta y_i , joka määritellään seuraavasti:

$$M_{i,j} = \{a_j, a_j+1, \dots, b_j-1, b_j\}, \quad (48)$$

$$\text{missä } b_j = \sum_{n=1}^j m_n \text{ ja } a_j = b_{j-1} + 1, \text{ kun } b_0 = 0.$$

Toisin sanoen kuukausi j alkaa indeksistä a_j ja päättyy indeksiin b_j , joka lasketaan summaamalla kuukausien $1 \dots j$ päivien lukumääriä, jotka on listattu vektorissa \mathbf{m} vuodelle y_i , joka ei ole karkausvuosi:

$$\mathbf{m}^T = (31 \ 28 \ 31 \ 30 \ 31 \ 30 \ 31 \ 31 \ 30 \ 31 \ 30 \ 31). \quad (49)$$

Jos vuosi y_i on karkausvuosi ja helmikuussa on 29 päivää, alkio m_2 on 29.

Jos esimerkiksi halutaan laskea maaliskuulle, eli $j = 3$, toisen vaiheen analyysit vuodelta, joka ei ole karkausvuosi, joukko $M_{i,3}$ olisi lukujono $\{60, 61, \dots, 89, 90\}$, jolla ehdon täyttävien indeksien X_i^{cond} ja pätevien indeksien J_i^x leikkausta leikattaisiin.

4.2.4 Pitkän aikavälin analyysi

Pitkän aikavälin analyysillä (*LTA – Long Term Analysis*) tarkoitetaan mittadatan analyysiä kaikilta niiltä kokonaisilta vuosilta, joilta mitta-asema on tuottanut mittauksia (ks. taulukko 7). Pitkän aikavälin analyysit perustuvat vuotuisten ja kuukausittaisten perusanalyysien sekä toisen vaiheen analyysien tuloksiin kaikilta kokonaisilta vuosilta aina viimeisimpään vuoteen (tässä työssä 2013).

Jos suurelle on laskettu jokin taulukon 8 perusanalyysi (keskiarvot, ääriarvot, kertymät), sille lasketaan myös pitkän aikavälin analyysi. Sen syötteenä on vuotuiset perusanalyysin tulokset. Mikäli suurelle on laskettu toisen vaiheen analyysijä, myös niistä lasketaan pitkän aikavälin analyysi, mutta vain vuotuisista tuloksista.

Olkoon vektori \mathbf{x}^Y jokin suurelle x laskettu vuotuinen perus- (kaavat 36, 37, 38, 41) tai toisen vaiheen analyysi (kaava 46) ja olkoon $_{DI}\mathbf{x}^Y$ niin ikään suurelle x laskettu mittadatan eheysvektori (kaava 43), jotka kattavat vuodet $y_1 \dots y_N$ ja sisältävät näin ollen N alkioita. Pitkän aikavälin analyysi tai keskiarvo x^{LTA} on:

$$x^{LTA} = \frac{\sum_{i=1}^N x_i^Y \theta({}_{DI}x_i^Y; \Theta)}{\sum_{i=1}^N \theta({}_{DI}x_i^Y; \Theta)}, \quad (50)$$

missä $\theta(x; \Theta)$ on määritelmän (44) mukainen funktio ja Θ mittadatan eheydelle asetettu kynnyсарvo. Samassa yhteydessä lasketaan analyysiin mukaan otettujen, luotettavien vuosien lukumäärä N^{LTA} , joka on kaavan (50) nimittäjä:

$$N^{LTA} = \sum_{i=1}^N \theta({}_{DI}x_i^Y; \Theta). \quad (51)$$

MetEnvissä pitkän aikavälin analyysin kynnyсарvoksi on asetettu 0,80 (tai 80%).

Edellä käsitellyn analyysin lisäksi kahdelle OL-WOM1 -asemalta mitatulle suurelle lasketaan pitkän aikavälin analyysi myös kuukausittaisten perusanalyysien pohjalta. Tämä analyysi on erikoistapaus ja se suoritetaan vain ilman lämpötilalle (maataso) sekä sademäärälle. Merkitään lämpötilaa jatkossa kirjaimella t .

Olkoon $\bar{\mathbf{T}}$ matriisi, jonka sarakkeet koostuvat suurelle t lasketuista kuukausittaisista keskiarvoista:

$$\bar{\mathbf{T}} = \left(\bar{\mathbf{t}}_1^M \quad \bar{\mathbf{t}}_2^M \quad \dots \quad \bar{\mathbf{t}}_j^M \quad \dots \quad \bar{\mathbf{t}}_{21}^M \right) \in \mathbb{R}^{12 \times 21}, \quad (52)$$

missä sarake j sisältää vuoden y_j kuukausittaiset keskiarvot, jotka on laskettu kaavalla (36) osana perusanalyysiä. Vuoden 2013 loppuun mennessä OL-WOM1 -asemalta on kertynyt mittadataa kaikkiaan 21:ltä vuodelta alkaen vuodesta 1993, eli $y_1 = 1993$ (ks. taulukko 7), mistä seuraa, että matriisissa $\bar{\mathbf{T}}$ on 21 saraketta.

Olkoon ${}_{DI}\mathbf{T}$ edellä kuvatun kaltainen matriisi, jonka sarakkeet koostuvat suurelle t lasketuista kuukausittaisista mittadatan eheyksistä:

$${}_{DI}\mathbf{T} = \left({}_{DI}\mathbf{t}_1^M \quad {}_{DI}\mathbf{t}_2^M \quad \dots \quad {}_{DI}\mathbf{t}_j^M \quad \dots \quad {}_{DI}\mathbf{t}_{21}^M \right) \in \mathbb{R}^{12 \times 21}, \quad (53)$$

missä sarakkeet ovat laskettu kaavalla (43) perusanalyysin yhteydessä.

Olkoon vektori $\bar{\mathbf{t}}^{LTAM}$ pitkän aikavälin kuukausittaiset keskiarvot:

$$\bar{\mathbf{t}}^{LTAM} \in \mathbb{R}^{12}, \text{ missä } t_i = \frac{\sum_{j=1}^{21} t_{i,j}^M \theta({}_{DI}t_{i,j}^M; \Theta)}{\sum_{j=1}^{21} \theta({}_{DI}t_{i,j}^M; \Theta)}. \quad (54)$$

Kaavassa (54) esiintyvät alkio t_{ij} ja ${}_{DI}t_{ij}$ ovat matriisien (52) ja (53) alkioita vastaavasti, sekä $\theta(t_{ij}; \Theta)$ määritelmän (44) mukainen funktio. MetEnvissä kynnyisarvo Θ on asetettu arvoon 0,90 (tai 90%).

Matriisista (52) etsitään myös suurimmat ja pienimmät kuukausittaiset keskiarvot, sekä niitä vastaavat indeksit, joiden perusteella voidaan määrittää vuosi. Olkoon vektori $\max \bar{\mathbf{t}}^{LTAM}$ suurimmat keskiarvot ja $\max \bar{\mathbf{j}}^{LTAM}$ niitä vastaavat indeksit:

$$\max \bar{\mathbf{t}}^{LTAM} \in \mathbb{R}^{12}, \text{ missä } t_i = \max_{j \in J_i} \bar{\mathbf{t}}_{i,*}^M \text{ ja} \quad (55)$$

$$\max \bar{\mathbf{j}}^{LTAM} \in \mathbb{N}^{12}, \text{ missä } j_i = \arg \max_{j \in J_i} \bar{\mathbf{t}}_{i,*}^M. \quad (56)$$

Joukko J_i kaavoissa (55) ja (56) on niiden indeksien joukko, joissa mittadatan eheys kuukaudelle i on vähintään yhtä suuri kuin asetettu kynnyksarvo:

$$J_i = \left\{ j \mid \forall j \in \{1, 2, \dots, 21\} : {}_{DI}t_{i,j}^M \geq \Theta \right\}, \quad i = 1, 2, \dots, 12, \quad (57)$$

missä Θ on kynnyksarvo, joka on edelleen 0,90, ja alkiot ${}_{DI}t_{i,j}$ matriisin (53) alkioita. Pienimmät keskiarvot sekä niiden indeksit saadaan samalla tavalla kaavoilla (55), (56) ja (57) käyttäen vastaavasti min ja arg min -funktioita.

Analyysin yhteydessä lasketaan myös niiden vuosien lukumäärä, joita ei ole otettu huomioon, koska mittadatan eheys on ollut liian pieni. Kaavan (54) nimittäjästä saadaan luotettavien vuosien lukumäärä. Kun se vähennetään kaikkien vuosien lukumäärästä, eli 21:stä, saadaan niiden vuosien määrä, jotka on jätetty väliin.

Kaikki operaatiot, joita tähän mennessä on tehty ilman lämpötilalle, ovat täysin analogisia sademäärälle, joten en käsittele sitä tässä sen yksityiskohtaisemmin. Sen sijaan ilman lämpötilalle etsitään vielä ääriarvojen ääriarvot. Tätä varten muodostetaan vielä kaksi matriisia $\max \mathbf{T}$ ja $\min \mathbf{T}$ samalla tavalla kuin matriisi (52):

$$\max \mathbf{T} = \begin{pmatrix} \max \mathbf{t}_1^M & \max \mathbf{t}_2^M & \dots & \max \mathbf{t}_j^M & \dots & \max \mathbf{t}_{21}^M \end{pmatrix} \in \mathbb{R}^{12 \times 21} \quad \text{ja} \quad (58)$$

$$\min \mathbf{T} = \begin{pmatrix} \min \mathbf{t}_1^M & \min \mathbf{t}_2^M & \dots & \min \mathbf{t}_j^M & \dots & \min \mathbf{t}_{21}^M \end{pmatrix} \in \mathbb{R}^{12 \times 21}, \quad (59)$$

missä sarake j sisältää vuoden y_j kuukausittaiset maksimit (58) tai minimi (59), jotka on laskettu vastaavasti kaavoilla (38) ja (41) perusanalyysin osana. Ääriarvojen ääriarvot saadaan yksinkertaisesti määritelmien (55) ja (56) funktioilla, vain syötteet ovat muuttuneet matriiseihin (58) ja (59). Mittadatan eheys on edelleen matriisi (53), jota tarvitaan joukon (57) muodostamiseen.

4.3 Vallitseva tuulen suunta

Vallitseva tuulen suunta (*PWD – Prevailing Wind Direction*) on analyysi, joka ilmaisee yleisimmän tuulen suunnan tietyltä jaksolta [16]. Tässä työssä PWD lasketaan joko pelkästään tuulen suunnan perusteella tai tuulen suunnan ja nopeuden perusteella. Analyysi lasketaan vain OL-WOM1 -asemalta mitatuille suureille, eli leviämissuunnalle ja tuulen nopeudelle kaikilta mittauskorkeuksilta (ks. taulukko 1). Suureista lasketaan kuitenkin päivittäiset, kuukausittaiset ja vuotuiset analyysit sekä vielä pitkän aikavälin analyysi.

Vallitseva tuulen suunta frekvenssin perusteella (*PWDF – PWD by Frequency*) on yksi menetelmä suorittaa tämä analyysi ja siinä hyödynnetään vain mittauksia tuulen suunnasta. Sen sijaan vallitseva tuulen suunta kertymän perusteella (*PWDA – PWD by Accumulation*) ottaa huomioon myös tuulen nopeuden. Jotta kumpaakaan menetelmää voidaan käyttää, mittadata (tuulen suunta) tulee ensin luokitella.

Tuulen suunta luokitellaan ilmansuuntien mukaan kahdeksaan 45° :n sektoriin siten, että ilmansuunnan kompassineula puolittaa sektorin. Luokat tai sektorit ovat puoliavoimia välejä. Esimerkiksi luokka 2, eli koillinen, käsittää välin $[22,5^\circ; 67,5^\circ)$. Taulukossa 12 esitetään nämä luokat. Tuulen suunta on validoinnin yhteydessä rajattu välille $[0^\circ; 360^\circ)$, kuten luvussa 3 on esitetty (ks. taulukko 4).

Taulukko 12: Luokat ilmansuuntien mukaan

Ilmansuunta	Luokka l	Väli I_l
pohjoinen	1	$I_1 = [0^\circ; 22,5^\circ) \cup [337,5^\circ; 360^\circ)$
koillinen	2	$I_2 = [22,5^\circ; 67,5^\circ)$
itä	3	$I_3 = [67,5^\circ; 112,5^\circ)$
kaakko	4	$I_4 = [112,5^\circ; 157,5^\circ)$
etelä	5	$I_5 = [157,5^\circ; 202,5^\circ)$
lounas	6	$I_6 = [202,5^\circ; 247,5^\circ)$
länsi	7	$I_7 = [247,5^\circ; 292,5^\circ)$
luode	8	$I_8 = [292,5^\circ; 337,5^\circ)$

Olkoon C taulukossa 12 esitettyjen luokkien (sektorien) joukko:

$$C = \{1, 2, 3, 4, 5, 6, 7, 8\} \subset \mathbb{N}. \quad (60)$$

Määritellään myös luokittelufunktio $c: [0^\circ; 360^\circ) \rightarrow \{0, 1\} \subset \mathbb{R}$, joka ottaa huomioon luokkaan l kuuluvan mittauksen:

$$c(\alpha; l) = \begin{cases} 1 & , \alpha \in I_l \\ 0 & , \alpha \notin I_l \end{cases}, \quad l \in C. \quad (61)$$

Luokittelijan argumentti on tuulen suunta (mittaus) ja I_l on taulukossa 12 esitetty joukko, joka valitaan parametrilla l .

Olkoon ϕ datavektori, joka sisältää mittaukset tuulen suunnasta, ja \mathbf{s}^ϕ vastaava statusvektori. Olkoon $\mathbf{F}^P \in \mathbb{Z}^{k \times 8}$ frekvenssimatriisi yleisille jaksoille $\mathbf{P}^{k \times 2}$ (ks. luku 4.1.1), jossa on k riviä ja 8 saraketta, eli yksi sarake kullekin taulukon 12 luokalle. Yläindeksi P merkitsee yhä analysoitavaa jaksoa, eikä sitä tule sekoittaa matriisiin \mathbf{P} . Matriisin \mathbf{F}^P alkiot lasketaan kaavan (61) avulla:

$$f_{i,l}^P = \sum_{j=\mathbf{P}(i,1)}^{\mathbf{P}(i,2)} c(\phi_j; l) v(s_j^\phi; V), \quad l \in C \quad \text{ja} \quad i = 1, 2, \dots, k, \quad (62)$$

missä V on määritelmän (5) joukko ja $v(s;V)$ määritelmän (6) funktio, sekä k analysoitavien jaksojen lukumäärä. Tämä analyysin vaihe on hyvin samankaltainen kuin mikä tahansa luvussa 4.2.1 kuvattu perusanalyysi, paitsi että tässä lasketaan yhden tunnusluvun sijaan frekvenssi jokaiselle luokalle l jaksolta i . Mittadatan eheys lasketaan täsmälleen kuten luvussa 4.2.2. on kuvattu, mutta DI-arvo kattaa aina kaikki frekvenssit jaksolta i sen sijaan, että eheys laskettaisiin erikseen jokaiselle luokalle.

Vallitseva tuulen suunta frekvenssin perusteella jaksolle P voidaan nyt määrittää frekvenssimatriisin avulla:

$$\mathbf{PWDF}^P \in C^k, \text{ missä } PWDF_i^P = \arg \max_{l \in C} \mathbf{f}_{i,*}^P, \quad i = 1, 2, \dots, k. \quad (63)$$

Funktion $\arg \max$ syöte on siis matriisin \mathbf{F}^P rivi i ja funktio palauttaa aina ensimmäisen indeksin, jossa maksimiarvo saavutetaan. Tämä indeksi vastaa suoraan jotakin taulukon 12 luokkaa, joka näin ollen on vallitseva tuulen suunta frekvenssin perusteella jaksolta i .

Olkoon \mathbf{w} datavektori, joka sisältää mittaukset tuulen nopeudesta, ja \mathbf{s}^w vastaava statusvektori. Olkoon $\mathbf{A}^P \in \mathbb{R}^{k \times 8}$ kertymämatriisi yleisille jaksolle $\mathbf{P}^{k \times 2}$, jonka alkiot lasketaan kaavan (61) avulla:

$$a_{i,l}^P = \sum_{j=\mathbf{P}(i,1)}^{\mathbf{P}(i,2)} w_j v(s_j^w; V) c(\phi_j; l) v(s_j^\phi; V), \quad l \in C \text{ ja } i = 1, 2, \dots, k, \quad (64)$$

missä V ja $v(s;V)$ ovat määritelmien (5) ja (6) mukaiset, sekä k analysoitavien jaksojen lukumäärä. Kaavalla (64) lasketaan siis tuulen nopeuksien kertymät kullekin luokalle l . Kaikki syötevektorit ovat keskenään synkronoituja, eli tuulen nopeutta w_j ja statusa s_j^w vastaa tuulen suunta ϕ_j ja staus s_j^ϕ . Kaavasta (64) havaitaan myös, että mittadata on keskenään validia, mistä seuraa, että mittadatan eheys lasketaan hieman aiemmasta poiketen. Eheyden laskemiseksi tässä erikoistapauksessa kaavan (43) osoittajaa tulee muokata huomioimaan kaksi statusvektoria:

$$\sum_{j=\mathbf{P}(i,1)}^{\mathbf{P}(i,2)} v(s_j^w; \{0\}) v(s_j^\phi; \{0\}) \quad (65)$$

Kaavan (43) nimittäjä ei muutu, koska jaksojen pituudet eivät riipu mittadatan oikeellisuudesta.

Kun kertymämatriisi on tiedossa, vallitseva tuulen suunta kertymien perusteella jaksolle P voidaan määrittää:

$$\mathbf{PWDA}^P \in C^k, \text{ missä } PWDA_i^P = \arg \max_{l \in C} \mathbf{a}_{i,*}^P, \quad i = 1, 2, \dots, k. \quad (66)$$

Funktion $\arg \max$ syöte on nyt matriisiin \mathbf{A}^P rivi i ja funktio palauttaa yhä ensimmäisen indeksin, jossa maksimiarvo saavutetaan, ja se vastaa suoraan taulukon 12 luokkaa, joka näin ollen on vallitseva tuulen suunta kertymän perusteella jaksolta i .

Kun vuotuiset analyysit on suoritettu ja vektorit \mathbf{PWDF}^Y ja \mathbf{PWDA}^Y , sekä niitä vastaavat mittadatan eheydet tiedetään, voidaan pitkän aikavälin analyysi laskea. Pitkän aikavälin vallitseva tuulen suunta saadaan selvittämällä edellä mainittujen vektorien moodit, eli yleisimmin esiintyvät arvot:

$$PWDF^{LTA} = Mo(\mathbf{PWDF}^Y) \quad \text{ja} \quad (67)$$

$$PWDA^{LTA} = Mo(\mathbf{PWDA}^Y), \quad (68)$$

missä funktio Mo palauttaa argumenttivektorin moodin. Analyysiin huomioitujen, luotettavien vuosien lukumäärä lasketaan kaavalla (51), jonka kynnsarvoksi on asetettu 0,80 (tai 80%). Kaavat (67) ja (68) myös olettavat, että argumenttivektoreista on poistettu ne alkio, joiden mittadatan eheys jää alle tämän kynnsarvon. Kaavan (67) tulos esitetään Posivan työraportissa [4].

Kappaleessa 2 kerrotaan, että leviämisseunta, eli suure, josta $PWDF$ sekä $PWDA$ lasketaan, on vastakkainen tuulen suunnalle. Toisin sanoen, jos leviämisseunta hetkeltä i on ϕ_i , tuulen suunta samalta hetkeltä i on silloin $\phi_i + 180^\circ$. Tässä kappaleessa kuvattu menetelmä olettaa, että suure on nimenomaan tuulen suunta, eikä leviämisseunta.

Tilanteen korjaamiseksi on kaksi tapaa. Ensimmäinen tapa on muokata mittadataa siten, että jokaiseen alkioon lisätään 180° ja ”täydet kierrokset” vähennetään alkioista. Helpompi tapa, joka ei vaadi datan muokkaamista, on muokata menetelmän tuloksia.

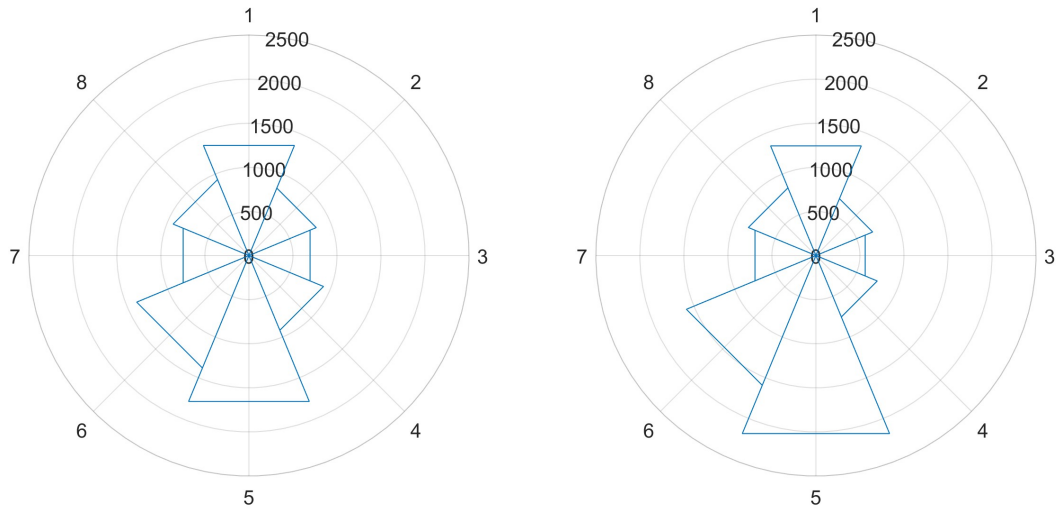
Olkoon PWD' joko frekvenssin tai kertymän perusteella laskettu vallitseva tuulen *leviämisseunta* miltä tahansa jaksolta tai pitkältä aikaväliltä, kun syöte on siis ollut leviämisseunta. Näin ollen vallitseva tuulen suunta PWD on:

$$PWD = \text{mod}(PWD' + 3 ; 8) + 1, \quad (69)$$

$$\text{missä } \text{mod}(x; N) = x - \text{floor}\left(\frac{x}{N}\right)N, \text{ kun } N \in \mathbb{N} \text{ ja } x > 0.$$

Funktio mod laskee argumentin x ja parametrin N osamäärän jakojäännöksen. Parametri on tässä taulukon 12 korkein luokka, eli 8.

Tuuliruusu on histogrammin kaltainen kuvaaja, josta ilmenee tuulen suunnan jakauma tietyssä paikassa ja tietyllä jaksolla. Se on oivallinen kuvaaja havainnollistamaan myös tässä kappaleessa kuvattujen menelmien eroa. Kuvassa 8 on esitetty tuulen suunnan jakaumat sekä frekvenssien että kertymien mukaan vuodelta 2013. Tuuliruusuksen puolet (kolmiot) vastaavat taulukon 12 luokkia. Kummastakin kuvaajasta käy selväksi, että vallitseva tuulen suunta on etelä (luokka 5).



Kuva 8: Vasemmalla tuulen suunta frekvenssin mukaan ja oikealla tuulen suunta kertymän mukaan OL-WOM1 -asemalta 20 metrin korkeudelta vuodelta 2013. Kaikki mittadata oli validia.

Koska OL-WOM1 -asemalla ei mitata tuulen suuntaa vaan leviämissuuntaa, dataa on muokattu kääntämällä leviämissuunta vastakkaiseksi, eli lisäämällä mittauksiin 180° ja poistamalla ”täydet kierrokset”. Vasemmanpuoleisessa jakaumassa kunkin kolmion korkeus vastaa vuotuisen frekvenssimatriisin \mathbf{F}^Y rivin $\mathbf{f}_{21,*}$ alkioita f_1, f_2, \dots, f_8 , joista siis alkio f_5 vastaa vallitsevaa tuulen suuntaa frekvenssien perusteella. OL-WOM1 -asemalta havaintovuosia on 21, joista viimeinen on vuosi 2013.

Oikeanpuoleisessa jakaumassa kunkin kolmion korkeus vastaa vuotuisen kertymämatriisin \mathbf{A}^Y rivin $\mathbf{a}_{21,*}$ alkioita a_1, a_2, \dots, a_8 siten, että jokainen alkio on suhteutettu vuoden kokonaiskertymään ja kerrottu validien havaintojen määrällä, jotta jakaumaa voidaan verrata frekvenssijakaumaan. Näin ollen luokan l skaalattu kertymä a_l' on:

$$a_l' = \frac{a_{21,l}}{\sum_{i=1}^8 a_{21,i}} \times 8760, \quad l \in C. \quad (70)$$

Koska kaikki mittaukset vuodelta 2013, joka ei ollut karkausvuosi, olivat valideja, havaintojen määrä on siten 8760. Menetelmien eron huomaa luokissa 5 ja 6, joissa puolat ovat selvästi pidempiä oikealla. Tämä tarkoittaa, että etelä- ja lounaistuulet ovat olleet myös voimakkaampia, eivät ainoastaan yleisempiä vuonna 2013.

4.4 Kasvukausi

Kasvukausi on se osa vuotta, jolloin kasvit kasvavat. Terminen kasvukausi (jatkossa pelkästään kasvukausi) määritetään vuorokautisten keskilämpötilojen perusteella.

Kasvukausi käsittää kaksi päivämäärää, jotka ilmaisevat kauden alun ja lopun. Nämä päivämäärät saadaan selville tiettyjen ehtojen täytyttyä. Ehtoja on useita ja ne poikkeavat toisistaan lähteestä riippuen.

Kasvukauteen liittyy läheisesti käsite *tehoisa lämpötila* (*effective temperature sum*). Posivan työraportin [6, s. 16] mukaan se lasketaan summaamalla kaikki vuorokautiset keskilämpötilat kasvukauden ajalta siten, että jokaisesta arvosta vähennetään ensin 5°C . Lisäksi vuorokaudet, joiden keskilämpötila on alle $+5^{\circ}\text{C}$ (ennen vähennystä), jätetään huomioimatta, ne eivät siis kasvata eikä vähennä summaa. Tehoisan lämpötilan yksikkönä käytetään vuorokausiastetta $^{\circ}\text{Cd}$.

Ilmatieteenlaitos määrittelee kasvukauden alkavan, ”kun lumi on sulanut aukeilta paikoilta ja vuorokauden keskilämpötila nousee pysyvästi yli $+5$ asteen. Koska keväällä vuorokauden keskilämpötila voi vaihdella pitkään $+5$ asteen molemmin puolin, tilannetta seurataan 10 vuorokauden ajan.” [14]

Edelleen ilmatieteenlaitos määrittelee kasvukauden päättyvän, ”kun vuorokauden keskilämpötila laskee pysyvästi $+5$ asteen alapuolelle tai saadaan pysyvä lumipeite. Syksyllä myös kovat yöpakkaset useana yönä peräkkäin katkaisevat kasvukauden.” [14]

Posivan työraportissa [6, s. 15] kasvukauden alku ja loppu määritetään konkreettisemmin. Kasvukausi alkaa, kun

1. vuorokauden keskilämpötila on yli $+5^{\circ}\text{C}$ vähintään viideltä peräkkäiseltä päivältä,
2. keskilämpötilojen summa näiltä viideltä päivältä on yli 20°Cd [sic], ja
3. aukeilla paikoilla ei ole lumipeitettä.

Kasvukausi päättyy, kun

1. vuorokauden minimilämpötila on alle -10°C , tai
2. vuorokauden keskilämpötila on alle $+5^{\circ}\text{C}$ vähintään viideltä peräkkäiseltä päivältä.

Alkuehto 2 on sekä virheellinen että redundantti. Ensinnäkin kynnysarvon 20°Cd yksikkö on väärä, kyseessä ei toki ole tehoisan lämpötilan summa vaan pelkkä keskimääräisten lämpötilojen summa viideltä peräkkäiseltä päivältä. Tätä käsitystä puoltaa sekä Ilmatieteenlaitos [14] että Ilmasto-opas [15]. Ehto on näin ollen redundantti, koska keskilämpötilojen summa viideltä päivältä, joista jokaisen keskilämpötila on vähintään $+5^{\circ}\text{C}$, on oltava ainakin 25°C , eli reilusti yli kynnysarvon.

Toteuttamani algoritmi, jolla kasvukausi määritetään, poikkeaa hieman näistä ehdoista: kasvukauden katsotaan alkavan, kun alkuehto 1 toteutuu, ja päättyvän, kun lopetusehto 2 toteutuu. Näin menetellään, koska MetEnv ei käsittele lumipeitteestä kerättyä dataa. On käyttäjän vastuulla tarkistaa, voiko MetEnvin ehdottama päivämäärä olla käypä päivämäärä lumipeitteen osalta. Käyttäjän tulee tarkistaa päivämäärät myös minimilämpötilan osalta. Tämä on ainoa analyysi, jota ei ole täysin automatisoitu ja vaatii siten käyttäjältä työtä. Syy tähän on ollut työkalun kehitysvaiheen epäselvyydet koskien kasvukauden määritteleviä ehtoja.

Kasvukausi määritetään ilman lämpötilan (maataso) päivittäisten keskiarvojen, jotka lasketaan perusanalyysin yhteydessä, perusteella ja se määritetään jokaiselle mitta-asemalle. Olkoon $\bar{t}^{i,D}$ tämä vektori vuodelta y_i . Algoritmi on kaksivaiheinen:

1. Etsitään vektorista $\bar{t}^{i,D}$ pienin indeksi, jonka osoittama alkio on *yli* $+5^\circ\text{C}$, ja tallennetaan tuo indeksi, olkoon se m' . Jos neljä seuraavaa alkioita ovat kaikki *yli* $+5^\circ\text{C}$, kasvukauden alkuindeksi m on tallennettu m' . Jos jokin näistä viidestä peräkkäisestä alkioista ei ylitä kynnyksarvoa, m' hylätään ja alkuindeksin etsimistä jatketaan edelleen. Olkoon löydetty alkuindeksi jatkossa m .
2. Alkaen kohdasta $m+5$, etsitään vektorista $\bar{t}^{i,D}$ indeksi, jonka osoittama alkio on *enintään* $+5^\circ\text{C}$ ja otetaan tuo indeksi talteen, olkoon se n' . Mikäli myös neljä seuraavaa alkioita ovat kaikki *enintään* $+5^\circ\text{C}$, kasvukauden loppuindeksi n on talteen otettu n' . Jos jokin neljästä, indeksiä n' seuraavasta alkioista kuitenkin ylittää kynnyksarvon, n' hylätään ja etsintää jatketaan edelleen. Olkoon löydetty loppuindeksi jatkossa n ($m < n$).

Kohta 2 sallii päivittäisen keskilämpötilan putoavan hetkellisesti alle kynnyksarvon, kuten kohta 1 sallii sen hetkellisesti nousevan yli kynnyksarvon. Algoritmi päättelee kasvukauden päivämäärät vain lämpötilan perusteella. Se ei ota huomioon mittadatan eheyttä, vaan eheydet ilmoitetaan kullekin päivämäärälle analyysin tulosten yhteydessä, jolloin luotettavuuden arviointi jää käyttäjälle.

Kasvukauteen liittyy myös muitakin tunnuslukuja kuin vain kauden alun ja lopun ajankohdat. Kasvukauden kesto lasketaan kumulatiivisena summana. Olkoon vektori \mathbf{d}^i kasvukauden kumulatiivinen kesto vuodelta y_i , jonka alkiot lasketaan:

$$d_j^i = \begin{cases} 1 & , j = m \\ d_{j-1}^i + 1 & , m < j \leq n \\ 0 & , j < m \vee j > n \end{cases} \quad , j = 1, 2, \dots, 365. \quad (71)$$

Myös tehoisa lämpötila lasketaan kumulatiivisena summana. Olkoon vektori ${}_{\text{EFF}}\mathbf{t}^i$ kumulatiivinen tehoisa lämpötila vuoden y_i kasvukaudelle, joka lasketaan keskilämpövektorin $\bar{t}^{i,D}$ alkioista:

$${}_{\text{EFF}}t_j^i = \begin{cases} \text{eff}(\bar{t}_j^{i,D}; +5^\circ\text{C}) & , j = m \\ {}_{\text{EFF}}t_{j-1}^i + \text{eff}(\bar{t}_j^{i,D}; +5^\circ\text{C}) & , m < j \leq n \\ 0 & , j < m \vee j > n \end{cases} \quad , j = 1, 2, \dots, 365, \quad (72)$$

$$\text{missä } \text{eff}(t; t_{\text{TH}}) = \begin{cases} t - t_{\text{TH}} & , t > t_{\text{TH}} \\ 0 & , t \leq t_{\text{TH}} \end{cases}.$$

Funktio $\text{eff}(t; t_{\text{TH}})$ palauttaa argumentin ”tehoisan osan”, eli kynnyksarvon t_{TH} ylittävän osan. Parametrin t_{TH} alaindeksi merkitsee kynnyksarvoa (*threshold*).

Kasvukaudelta lasketaan myös kauden sademäärä, joka sekkin on kumulatiivinen summa. Olkoon vektori ${}_{\text{GSP}}\mathbf{p}^i$ kasvukauden kumulatiivinen sademäärä (*GSP – Growth Season Precipitation*) ja vektori ${}_{\Sigma}\mathbf{p}^{i,D}$ perusanalyysissä laskettu päivittäiset sademäärät, molemmat vuodelta y_i . Vektorin ${}_{\text{GSP}}\mathbf{p}^i$ alkiot lasketaan:

$${}_{\text{GSP}}\mathbf{p}_j^i = \begin{cases} {}_{\Sigma}P_j^{i,D} & , j = m \\ {}_{\text{GSP}}\mathbf{p}_{j-1}^i + {}_{\Sigma}P_j^{i,D} & , m < j \leq n \\ 0 & , j < m \vee j > n \end{cases} , j = 1, 2, \dots, 365. \quad (73)$$

Kaavoissa (71), (72) ja (73) oletetaan, että vuosi y_i ei ole karkausvuosi, muutoin indeksin j maksimi olisi tietysti 366. Varsinaiset kasvukauden tunnusluvut, eli kesto, tehoisa lämpötila ja kauden sademäärä löytyvät edellä määritellyistä vektoreista kohdasta, jossa kasvukausi päättyy, eli indeksin n osoittamasta kohdasta. Sademäärää ei mitata OL-WOM5 -asemalla, joten kasvukauden sademäärää ei lasketa kyseiseltä asemalta.

MetEnv esittää käyttäjälle kasvukauden aukaistuna pelkkien tunnuslukujen sijaan. Käytännössä tämä tarkoittaa taulukkoa, jossa esitetään vuoden y_i päivämäärät, päivittäiset keskilämpötilat, kumulatiivinen kesto, kumulatiivinen tehoisa lämpötila, keskilämpötilojen päivittäiset eheydet, kumulatiivinen sademäärä sekä päivittäisten sademäärien eheydet. Taulukosta käyttäjän on helppo tarkistaa, onko avoimilla paikoilla vielä lumipeite kasvukauden alussa, ja lukea tarvittavat tunnusluvut.

Työraporttia [4] varten pitää vielä määrittää pitkän aikavälin tunnusluvut kasvukausilta. Ne ovat kauden keskimääräinen alku- ja loppuajankohta, aikaisin ja myöhäisin alku- ja loppuajankohta, keskimääräinen kesto, lyhyin ja pisin kesto, keskimääräinen tehoisa lämpötila, pienin ja suurin tehoista lämpötila, keskimääräinen sademäärä, sekä pienin ja suurin sademäärä. MetEnviin näitä ei ole kuitenkaan toteutettu kehitysvaiheen epäselvyyksistä johtuen, joten myös pitkän aikavälin analyysi on kasvukauden osalta jätetty käyttäjän vastuulle.

5 YHTEENVETO

Posivan toimeksiannosta on toteutettu uusi versio MetEnv-työkalusta, jonka käyttö on aikaisempaan verrattuna vaivattomampaa, sekä tarkoituksenmukaisempaa. Työkalun käytössä tähän mennessä ilmenneet virheetkin on korjattu. Tässä diplomityössä on keskitytty kuvaamaan työkaluun liittyviä keskeisiä osa-alueita, kuten havaintoaineiston validointia ja mittadatan korjausmenetelmiä, sekä aineiston analyysiä. Diplomityössä on viitattu myös lähteisiin, joita on käytetty työkalun toteutustyön spesifikaatioina siitä, miten aineistoja tulee käsitellä.

Havaintoaineistosta kyetään automaattisesti tunnistamaan ja merkitsemään puuttuvat sekä epäilyttävän pienet ja suuret arvot. Tässä työssä on kuvattu kolme menetelmää, vakiolla täyttö, interpolointi ja profilointi, joilla puuttuvista mittauksista johtuvat aukot aineistossa voidaan korjata. Näistä menetelmistä suositellaan käytettävän profilointia, jos vain mahdollista, mutta käyttäjä tekee kuitenkin lopullisen päätöksen kunkin menetelmän käytöstä. Eri menetelmiä on myös arvioitu niiden aiheuttaman virheen näkökulmasta ja käyttäjälle on annettu ohjeistuksia mitä menetelmää milloinkin kannattaa soveltaa.

Suurin osa diplomityöstä kuvaa havaintoaineiston analyysiä. Validi aineisto analysoidaan neljässä kokonaisuudessa. Ensin suoritetaan perusanalyysi, jossa lasketaan keskiarvoja, kertymiä ja ääriarvoja päivittäisille, kuukausittaisille ja vuotuisille jaksoille. Toisessa vaiheessa lasketaan päivittäisen perusanalyysin pohjalta tietyt kriteerit täyttävien päivien lukumääriä. Tämän jälkeen lasketaan pitkän aikavälin analyysi, jossa hyödynnetään perusanalyysin vuotuisia sekä kuukausittaisia tuloksia. Lopuksi määritetään vielä kasvukausi. Analyysiin on toteutettu myös mittadatan eheyttä kuvaava tunnusluku, jonka perusteella voidaan arvoida analyysin luotettavuutta. Tämä tunnusluku vaikuttaa myös perusanalyysin jälkeisiin analyyseihin siten, että sen parusteella määritellään, mitkä analyysin tuloksista kelpuutetaan näihin jatko-analyyseihin.

Itse käyttöliittymään tai muuhun ohjelmistokehitykseen en ole diplomityössä juurikaan pureutunut, joten en lähde arvioimaan jatkokehitystyön tarvetta näiden kannalta. Sen sijaan kasvukauden määrittäminen on osa-alue, joka on jäänyt kesken ja se voisi olla aiheellista saattaa valmiiksi. Tämä vaatii ainakin kasvukauden määrittämisen lukkoonlyömistä, ja kun tuo spesifikaatio on selvillä, analyysi olisi helppo saattaa valmiiksi. Sen yhteyteen tulisi myös toteuttaa kasvukauden pitkän aikavälin analyysikin. Toinen jatkokehitystyön kohde voisi olla korjausmenetelmät. Profilointia voisi laajentaa hyödyntämään mittadataa myös muilta mitta-aseteilta, varsinkin asemilta

2, 3 ja 4, jotka sijaitsevat kaikki lähellä toisiaan. Myös polynomisia interpolointeja voisi toteuttaa, kuten neliöllinen ja kuutiollinen interpolointi.

MetEnv-työkalulle asetetut vaatimukset sen käyttötapauksen, monitorointiraporttien tarpeiden sekä Posivan asiantuntijoiden tahoilta ovat toteutuneet edellä mainitut jatkokehityskohteet huomioiden hyvin. MetEnvillä on analysoitu vuoden 2013 säähavaintoaineisto työkalun vielä ollessa kehitysvaiheessa. Analyysin tulokset on myöhemmin julkaistu Posivan ympäristön monitorointiraportissa [4]. Tämän lisäksi työkalun valmiilla versiolla on onnistuneesti analysoitu sekä vuosien 2014 että 2015 säähavaintoaineisto.

LÄHTEET

- [1] Monitoring at Olkiluoto – a Programme for the Period Before Repository Opening, Posiva 2012-01, Posiva, 2012.
Saatavissa (23.10.2016): www.posiva.fi/en/databank/posiva_reports
- [2] Programme of Monitoring at Olkiluoto During Construction and Operation of the ONKALO, Posiva 2003-05, Posiva, 2003.
Saatavissa (23.10.2016): www.posiva.fi/en/databank/posiva_reports
- [3] L. Aro, A-J. Lindroos, P. Rautio, Results of Forest Monitoring on Olkiluoto Island in 2013, Posiva Working Report 2015-06, Posiva, 2015.
Saatavissa (23.10.2016): www.posiva.fi/en/databank/working_reports
- [4] T. Pere, L. Aro, M. Tuohimaa, Results of Monitoring at Olkiluoto in 2013, Environment, Posiva Working Report 2014-45, Posiva, 2015.
Saatavissa (23.10.2016): www.posiva.fi/en/databank/working_reports
- [5] A. Haapanen, Results of Monitoring at Olkiluoto in 2012, Environment, Posiva Working Report 2013-45, Posiva, 2014.
Saatavissa (23.10.2016): www.posiva.fi/en/databank/working_reports
- [6] A. Ikonen, Meteorological Data and Update of Climate Statistics of Olkiluoto 2005–2006, Posiva Working Report 2007-86, Posiva, 2007.
Saatavissa (23.10.2016): www.posiva.fi/en/databank/working_reports
- [7] Säteilyturvakeskus (STUK) ohje YVL C.4 / 17.3.2015 – Ydinlaitoksen ympäristön väestön säteilyannosten arviointi, 3. luku.
Saatavissa (23.10.2016):
plus.edilex.fi/stuklex/fi/lainsaadanto/saannosto/YVLC-4/
- [8] ICP Forests – International Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests, verkkosivu.
Saatavissa (23.10.2016): icp-forests.net/

- [9] H. Raitio, A. Hamari, P. Merilä, J. Mäkinen, Yleiseurooppalainen metsien terveydentilan seuranta (YK-ECE/EK) Suomessa.
- [10] S. Raspe, E. Beuker, T. Preushler, A. Bastrup-Birk, Manual on methods and criteria for harmonized sampling, assessment, monitoring and analysis of the effects of air pollution on forests – part IX Meteorological Measurements, Hampuri, Saksa, 2010.
Saatavissa (23.10.2016): icp-forests.net/page/icp-forests-manual
- [11] M. Ferretti, R. Fischer, Developments in Environmental Science, vol 12: Forest Monitoring – Methods for terrestrial investigations in Europe with an overview of North America and Asia, Elsevier, 2013.
- [12] WebMET.com – The Meteorological Resource Center, verkkosivu.
Saatavissa (23.10.2016): www.webmet.com/met_monitoring/681.html
- [13] R. A. Adams, C. Essex, Calculus – a Complete Course, Pearson Canada Inc., Toronto, Ontario, Canada, 2010.
- [14] Ilmatieteenlaitos, verkkosivu.
Saatavissa (23.10.2016): ilmatieteenlaitos.fi/terminen-kasvukausi
- [15] Ilmasto-opas.fi, verkkosivu.
Saatavissa (23.10.2016): ilmasto-opas.fi/fi/
- [16] American Meteorological Society – Glossary of Meteorology, verkkosivu.
Saatavissa (23.10.2016): glossary.ametsoc.org/wiki/Main_Page
- [17] T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, P.M. Midgley, Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change – Annex III: Glossary, Cambridge University Press, Cambridge, Yhdistynyt kuningaskunta ja New York, Yhdysvallat, 2013.
Saatavissa (23.10.2016): www.climatechange2013.org/report/full-report/

LIITE A – OL-WOM2 -ASEMALTA MITATTAVAT SUUREET

Alla olevassa taulukossa on kaikki OL-WOM2 -asemalta mitattavat suureet. Joillekin suureille kirjataan keskiarvon lisäksi tunnitaiset minimi ja maksimit, joista jokainen muodostaa oman suureensa. Sademäärästä kirjataan tunnin kertymä keskiarvon sijaan.

Maan lämpötilaa mitataan 10 cm:n välein yhdeksältä eri syvyydeltä. Näistä 10 ja 20 cm:n syvyyttä mitataan kolmella anturilla kumpaakin muodostaen yhteensä 6 suuretta. Maan kosteutta mitataan kahdella anturilla, joten suureita on myös 2.

Kahteen puuhun on asennettu kasvupannat, joiden lukemat kirjataan tunneittain. Tämän lisäksi kirjataan myös akun jännite tunneittain.

Suure [yksikkö]	Mittaus- korkeus [m]	Anturien lkm	Keskiarvo	Minimi	Maksimi
Ilman lämpötila [°C]	2	1	X		
Ilman lämpötila [°C]	9	1	X	X	X
Ilman lämpötila [°C]	24	1	X	X	X
Ilman suhteellinen kosteus [%]	2	1	X		
Ilman suhteellinen kosteus [%]	9	1	X	X	X
Ilmanpaine [hPa]	2	1	X		
Maan lämpötila [°C]	-0,1	3	X		
Maan lämpötila [°C]	-0,2	3	X		
Maan lämpötila [°C]	-0,3	1	X		
Maan lämpötila [°C]	-0,4	1	X		
Maan lämpötila [°C]	-0,5	1	X		
Maan lämpötila [°C]	-0,6	1	X		
Maan lämpötila [°C]	-0,7	1	X		
Maan lämpötila [°C]	-0,8	1	X		
Maan lämpötila [°C]	-0,9	1	X		
Maan kosteus [%]	-0,2	2	X		
Tuulen nopeus [m/s]	24	1	X	X	X
Tuulen suunta [°]	24	1	X	X	X
Sademäärä [mm]	1	1			
Kokonaissäteily [W/m ²]	24	1	X	X	X
PAR [μmol/s/m ²]	24	1	X	X	X

Suure [yksikkö]	Mittaus- korkeus [m]	Anturien lkm	Keskiarvo	Minimi	Maksimi
Kasvupanta, puu nro 395 [mm]	-	1			
Kasvupanta, puu nro 93 [mm]	-	1			
Akun jännite [V]	-	1			

LIITE B – MITTADATATIEDOSTOESIMERKIT

Alla on osa tiedostoa Olkiluoto_kanava07.csv (rivit 3889 – 3914) asemalta OL-WOM2 vuodelta 2013. Tiedosto sisältää mittauksia maaperän lämpötilasta 90 cm:n syvyydestä. Siinä on 8760 riviä, joista jokaisella on mittaus. Mittadata on siis täydellinen, eikä kaipaa paikkaamista.

```

12.06.2013;01:00;6,66
12.06.2013;02:00;6,64
12.06.2013;03:00;6,66
12.06.2013;04:00;6,66
12.06.2013;05:00;6,66
12.06.2013;06:00;6,66
12.06.2013;07:00;6,66
12.06.2013;08:00;6,66
12.06.2013;09:00;6,66
12.06.2013;10:00;6,66
12.06.2013;11:00;6,68
12.06.2013;12:00;6,66
12.06.2013;13:00;6,66
12.06.2013;14:00;6,68
12.06.2013;15:00;6,68
12.06.2013;16:00;6,68
12.06.2013;17:00;6,68
12.06.2013;18:00;6,68
12.06.2013;19:00;6,68
12.06.2013;20:00;6,68
12.06.2013;21:00;6,68
12.06.2013;22:00;6,68
12.06.2013;23:00;6,70
13.06.2013;00:00;6,68
13.06.2013;01:00;6,70
13.06.2013;02:00;6,70

```

Tiedoston säännönmukaisuus on ilmiselvä: päivämäärä ja tunti muodostavat aikaleiman, jota varsinainen mittaus seuraa. Kaikki osat erotetaan puolipisteillä. Mitta-asemilta OL-WOM2, OL-WOM3, OL-WOM4 ja OL-WOM5 saatavat csv-mittadatatieostot noudattavat tätä yksinkertaista muotoa.

Seuraavalla sivulla on puolestaan 22 ensimmäistä riviä tiedostosta uuze01.13 asemalta OL-WOM1 tammikuulta 2013 (data on uuze-tiedostoissa kuukausittaista). Rivejä on kaikkiaan 746, joista ensimmäinen sisältää antureiden tunnisteen, toinen mittayksiköt ja loput 744 varsinaiset mittaukset. Nyt osa mittauksista näyttää uupuvan. Ilman suhteellinen kosteus (K811) on vailla yhtään mittausta ja kaksi mittausta puuttuu myös stabiilisuusluokan (K011) alusta.

DD.MM.YY	HH	K011	K501	K561	K521	K511	K001	K061	K021	K805	K865	K825	K801	K861	K821	K311	K911	K111	K811
			C	C	C	C	DEG	DEG	DEG	DEG	DEG	DEG	M/S	M/S	M/S	MM	MIN	hPA	% RH
01.01.13	00	**	2.34	2.54	2.69	2.62	4.7	6.5	13.7	343.	335.	335.	11.0	8.7	5.2	0.1	54.	988.	****
01.01.13	01	**	2.63	2.85	3.03	2.91	5.7	6.7	13.5	349.	340.	339.	10.5	8.5	5.3	0.6	59.	987.	****
01.01.13	02	5.	2.99	3.17	3.41	3.15	6.6	8.7	14.1	353.	350.	350.	9.3	7.1	4.5	0.6	57.	987.	****
01.01.13	03	5.	3.58	3.76	3.98	3.66	5.7	8.1	12.4	19.	2.	356.	10.8	8.4	5.7	0.5	57.	986.	****
01.01.13	04	5.	3.66	3.81	4.09	3.77	5.8	8.3	12.3	29.	19.	12.	12.9	9.9	6.9	0.3	57.	986.	****
01.01.13	05	4.	3.51	3.64	3.94	3.71	5.6	8.2	12.6	28.	18.	11.	12.8	9.9	6.8	1.6	55.	986.	****
01.01.13	06	4.	3.61	3.83	4.11	3.77	5.4	7.8	12.3	26.	17.	9.	12.6	9.9	6.8	1.7	60.	985.	****
01.01.13	07	4.	3.65	3.88	4.17	3.88	5.7	7.9	12.3	27.	17.	9.	12.5	9.6	6.6	2.0	60.	985.	****
01.01.13	08	4.	3.56	3.78	4.08	3.79	6.0	7.7	12.3	29.	19.	11.	12.4	9.7	6.7	1.8	60.	985.	****
01.01.13	09	4.	3.53	3.73	4.03	3.72	5.5	7.5	12.7	30.	20.	13.	12.2	9.6	6.5	1.8	60.	985.	****
01.01.13	10	4.	3.60	3.77	4.07	3.80	6.0	8.0	13.6	32.	23.	16.	13.0	9.9	6.8	0.3	27.	985.	****
01.01.13	11	4.	3.83	4.01	4.27	3.93	5.8	8.1	12.5	31.	21.	13.	14.3	11.1	7.7	.0	3.	985.	****
01.01.13	12	4.	3.64	3.85	4.19	3.93	6.2	8.3	13.8	32.	22.	14.	13.9	10.7	7.2	0.3	48.	986.	****
01.01.13	13	4.	3.44	3.59	3.92	3.71	5.8	7.8	13.5	35.	26.	17.	14.1	10.7	7.2	0.2	49.	986.	****
01.01.13	14	4.	3.17	3.35	3.71	3.52	5.8	7.9	14.4	39.	30.	21.	13.8	10.4	6.7	0.1	33.	987.	****
01.01.13	15	4.	2.74	2.94	3.27	3.10	5.4	7.6	13.5	37.	27.	20.	13.8	10.4	6.9	0.2	36.	987.	****
01.01.13	16	4.	2.56	2.80	3.11	2.84	5.8	7.5	13.5	36.	26.	16.	12.7	9.6	6.2	.0	4.	987.	****
01.01.13	17	4.	2.45	2.64	2.96	2.74	6.0	7.8	14.3	40.	31.	23.	13.2	9.7	6.2	.0	10.	988.	****
01.01.13	18	4.	2.49	2.71	3.06	2.83	5.5	7.8	13.8	40.	32.	25.	12.9	9.6	6.3	.0	3.	989.	****
01.01.13	19	4.	2.24	2.44	2.81	2.65	5.1	8.0	13.6	41.	32.	24.	12.1	8.9	5.7	.0	0.	989.	****

Säännönmukaisuus on edelleen selvä: päivämäärä ja tunti, joita seuraa mittaukset välilyönneillä erotettuna. Nämä uuze-tiedostot sisältävät aina kaksi otsikkoriviä: ensimmäinen sisältää antureiden tunnisteen ja toinen mittayksiköt, kuten aiemmin totesin. OL-WOM1 -aseman mittadata on aina tässä taulukkomuodossa, joka on ehkäpä tarpeettomankin monimutkainen verrattuna csv-tiedostoihin.