



OPERATIIVSETE ILMAMUDELITE ANALÜÜS JA VÕRDLUS TALLINNA LAHEL

Bakalaureusetöö

Üliõpilane: Oskar Villem Tiidemann

Üliõpilaskood: YAFB222492

Juhendaja: Aarne Männik, vanemteadur, TalTech meresüsteemide instituut

Õppekava: Rakendusfüüsika

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Oskar Villem Tiidemann

23.05.2025

Annotatsioon

Töö keskendub ECMWF ja MEPS ilmaennustusmudelite võrdlusele Tallinna lahe piirkonnas ning masinõppel põhinevate korrektsioonialgoritmide väljatöötamisele nende mudelite täpsuse parandamiseks. Töö eesmärgiks oli hinnata kahe operatiivse ilnamudeli – Euroopa Keskpika Ilmaennustuste Keskuse (ECMWF) globaalse mudeli ja METCoOp ansambelennustussüsteemi (MEPS) kontrollmudeli – prognoosikvaliteeti Tallinna lähel. Uurimistöö käigus analüüsiti nelja ilmajaama (Pirita, Tallinn-Harku, Rohuneeme, Naissaare) andmeid perioodist 2014-2024, võrreldes neid ECMWF ja MEPS mudelite ennustustega. Tulemused näitasid, et mõlemal mudelil on omad tugevused ja nõrkused. ECMWF mudel oli täpsem temperatuuri ja tuule suuna ennustamisel, samas kui MEPS mudel oli parem tuule kiiruse prognoosis. Mõlemad mudelid näitasid süstemaatilisi kõrvalekaldeid: ECMWF alahindas temperatuuri ja nihutas tuule suunda päripäeva, MEPS ülehindas temperatuuri ja nihutas tuule suunda vastupäeva. Süvaõppe meetoditel põhinev korrektsioonimudel rakendati mõlema ilnamudeli tulemuste täpsustamiseks. Kasutati mitmekihilist närvivõrku, mis võttis arvesse originaalseid mudeli ennustustusi, ajalisi andmeid ja hooajalisi indikaatoreid. Masinõppe algoritm parandas mõlema mudeli prognoose kõigis parameetrites. Lisaks töötati välja üllähedaste ennustuste mudel, mis kasutas eelnevate tundide mõõdetud andmeid järgneva tunni prognoosi täpsustamiseks. See lahendus andis märkimisväärselt paremaid tulemusi. Töö tulemused kinnitavad, et masinõppel põhinevad korrektsioonialgoritmid suudavad tõhusalt parandada traditsiooniliste numbriliste ilnamudelite täpsust piirkondlikes tingimustes.

Abstract

Analysis and comparison of operative weather prediction models in Tallinn bay

The work focuses on the comparison of ECMWF and MEPS weather forecast models in the Tallinn Bay region and the development of correction algorithms based on machine learning to improve the accuracy of these models. The aim of the work was to assess the forecast quality of two operational weather models – the European Centre for Medium-Range Weather Forecasts (ECMWF) global model and the METCoOp ensemble prediction system (MEPS) control model – in Tallinn Bay. Data was analyzed from four weather stations (Pirita, Tallinn-Harku, Rohuneeme, Naissaare) for the period 2014-2024. It was compared with the predictions of the ECMWF and MEPS models. The results showed that both models have their strengths and weaknesses. The ECMWF model was more accurate in predicting temperature and wind direction, while the MEPS model was better in predicting wind speed. Both models showed systematic deviations: ECMWF underestimated the temperature and shifted the wind direction clockwise, MEPS overestimated the temperature and shifted the wind direction counterclockwise. A correction model based on deep learning methods was applied to refine the results of both weather models. A multilayer neural network was used, which took into account the original model predictions, temporal data and seasonal indicators. The machine learning algorithm improved the forecasts of both models in all parameters. In addition, a nowcasting model was developed, which used measured data from previous hours to refine the forecast for the next hour. This solution gave significantly better results. The results of the work confirm that correction algorithms based on machine learning can effectively improve the accuracy of traditional numerical weather models in regional conditions.

Lühendite ja mõistete sõnastik

BIAS	Süsteemaatiline viga, kõrvalekalle (<i>Systematic Error, Bias</i>)
ECMWF	Euroopa Keskpika Ilmaennustuste Keskus (<i>European Centre for Medium-Range Weather Forecasts</i>)
ECWAM	ECMWF lainete mudel (<i>ECMWF Wave Model</i>)
HARMONIE-AROME	Kõrglahutusega regionaalne ilmamudel (<i>Hirlam Aladin Research towards Mesoscale Operational NWP in Europe - AROME</i>)
HTESSEL	Maismaa pinna hüdroloogia mudel (<i>Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land</i>)
IFS	Integreeritud prognoosimudel (<i>Integrated Forecasting System</i>)
LAM	Piiratud ala mudel (<i>Limited Area Model</i>)
LIM	Merejää mudel (<i>Louvain-la-Neuve Sea Ice Model</i>)
MAE	Keskmine absoluutne viga (<i>Mean Absolute Error</i>)
MARS	Meteoroloogiliste arhiivide ja otsimise süsteem (<i>Meteorological Archival and Retrieval System</i>)
MEPS	METCoOp ansamblennustussüsteem (<i>MetCoOp Ensemble Prediction System</i>)
METCoOp	Meteoroloogiline koostöö operatiivses numbrilises ilmaennustuses (<i>Meteorological Co-operation on Operational Numerical Weather Prediction</i>)
ML	Masinõpe (<i>Machine Learning</i>)
NEMO	Ookeanimudel (<i>Nucleus for European Modelling of the Ocean</i>)
NWP	Numbriline ilmaennustus (<i>Numerical Weather Prediction</i>)
RMSD	Ruutkeskmine erinevus (<i>Root Mean Square Difference</i>)

Sisukord

1 Sissejuhatus	6
2 Kirjanduse ülevaade	7
2.1 Numbrilised ilmaennustuse mudelid	7
2.1.1 Numbrilise ilmaennustuse põhiprintsiibid	7
2.1.2 Numbriliste meetodite piirangud	7
2.1.3 ECMWF mudeli ülevaade	8
2.1.4 MEPS mudeli ülevaade ja eripärad	8
2.2 Masinõpe ilmaennustuses: evolutsioon ja rakendused	9
2.2.1 Ajalooline areng ja praegune roll	9
2.2.2 Algoritmid, koolitusmeetodid ja väljakutsed	9
2.2.3 Korrektsioonimudelid ja arvutuslike väljakutsete lahendamine	10
3 Metoodika	11
3.1 Andmete kogumine	11
3.2 Võrdlusmetoodika	12
3.3 Korrektsioonimudel	13
3.3.1 Sissejuhatus	13
3.3.2 Mudeli arhitektuur ja disain	13
3.3.3 Parameetrite valik ja kalibreerimine	13
3.3.4 Hetkennustuste mudel	14
4 Tulemused	15
4.1 Esialgsete ilmaennustusmudelite võrdlus	15
4.2 Korrektsioonimudeli tulemused	18
4.3 Hetkennustuse võrdlus	20
4.4 Ennustuste kokkuvõte	22
5 Järeldused ja arutelu	24
6 Kokkuvõte	26
Kasutatud kirjandus	27
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	30

1. Sissejuhatus

Ilmaennustus on kriitilise tähtsusega valdkond, mis mõjutab valdkondi põllumajandusest kuni merenduseni välja. Tänapäeval muutub üha olulisemaks võime ennustada ilma võimalikult täpselt ning kohalikke eripärasid arvestades. Globaalsed ilnamudelid ei suuda tihti kirjeldada kohalikke tingimusi väiksematel piirkondadel, näiteks lahtedel piisava täpsusega.

Läänemere suur ja keeruka kujuga veekogu avaldab olulist mõju kohalikule kliimale ning põhjustab suurt varieeruvust pinnatuultes kogu basseini ulatuses. Täpsemalt, Soome lahe tuulerežiim peegeldab mitmeid basseini-spetsiifilisi või piirkihiefekte, kombineerides domineerivaid edelatuuli ja põhjatuuli kohalike ida- ja läänetuultega, mis puhuvad piki lahe telge.

Soome lahe keskosa, sealhulgas Tallinna lahe piirkond, on tiheda laevaliiklusega piirkond, kus ristuvad kaks olulist laevateed: Tallinn-Helsingi reisiliinid ja Soome lahe telje suunaline laevatee. Selles piirkonnas toimuvad ka tihti võistluspurjetamise etapid. Usaldusväärne meteoroloogiline teave on seega kriitilise tähtsusega nii navigatsiooniriskide ja lainetuskoormuste hindamisel kui meretranspordi, rannikuala turismi ja kohaliku elanikkonna ohutuse tagamisel. Globaalsed ilnamudelid ei suuda tihti piisava täpsusega arvestada kohalike eripäradega, seega on vajalik hinnata ja analüüsida erinevate operatiivsete mudelite kvaliteeti konkreetses piirkonnas.

Töö eesmärgiks on võrrelda ECMWF(European Centre for Medium range Weathere Forecasts) ja MEPS_{control} (METCoOP Ensemble Prediction System) mudelid Tallinna lahe piirkonnas. Peale selle, on eesmärgiks testida masinõppepõhiste korrektsioonialgoritmide tulemust mõlemale mudelile samas piirkonnas.

Analüüsimisel keskendutakse mitmele ilmajaamale, mis asuvad erinevates strateegilistes asukohtades, need olid Harku, kus tuuleandmed on sageli vähese seosega avamere tuulerežiimiga, Piritasadam, mis teatud tuulesuundade ja -kiiruste korral esindab merel valitsevaid tuuli paremini, Naissaare lõunaosas asuv mõõtejaam, mis esindab merelisi olustike ning Rohuneemel asuv mõõtejaam.

2. Kirjanduse ülevaade

2.1 Numbrilised ilmaennustuse mudelid

2.1.1 Numbrilise ilmaennustuse põhiprintsiibid

Valdav enamus ilmaennustusest käib tänapäeval läbi numbriliste meetodite. Numbrilise ilmaprognoosi mudelid kasutavad arvutuslikke meetodeid atmosfääri käitumise modelleerimiseks ja ennustamiseks. Numbrilised ilmaprognosid lahendavad atmosfääri liikumist kirjeldavaid diferentsiaalvõrrandeid, jagades atmosfääri kolmemõõtmeliseks võrgustikuks, kus igal punktil on oma temperatuur, rõhk, niiskus ja tuuleparameeter [1]. Numbrilise ennustusviisi tööpõhimõte on järgmine: ilmaennustus teisendab atmosfääris toimuva füüsika matemaatilisteks võrranditeks, mida suudavad lahendada arvutid. Esiteks jaotatakse atmosfäär kolmemõõtmeliseks võrguks. Siis määratakse igale punktile algsed väärtused vastavalt mõõtmisandmetele. Peale seda rakendatakse füüsikalisi seaduseid, et arvutada kuidas need väärtused muutuvad ajas. Saadud tulemust kasutatakse järgmise simulatsiooni astme algandmetena. Nende simulatsioonide täpsus sõltub sellest, kui suur on võrgustiku eraldusvõime. Mida väiksemateks ruumihulkadeks jagatakse seda täpsem simulatsioon, aga seda rohkem on tarvis arvutusvõimsust.

Numbrilise ilmaennustusmudelite oluline aspekt on ääritingimuste käsitlemine. Piiratud ala mudelites (LAM-ides), näiteks MEPS-I, on külgmised ääritingimused olulised, kuna need määravad kuidas liigub info suuremast (globaalsest) mudelist regionaalsesse mudelisse [1]. Ääritingimustega seotud vead võivad levida mudeli sisemusse ja nende mõju vähendamiseks kasutatakse "puhvervööndite" tehnikat.

Kaasaegsed operatiivmudelid, nagu ECMWF globaalne mudel ja MEPS regionaalne süsteem, on tänapäevase numbrilise prognoosimudeli rakendamise kõige arenenumad näited. Nende mudelite kõrge lahutusvõime, keerukas füüsika ja andmete kaasamise tehnoloogiad võimaldavad modelleerida atmosfääri protsesse väga täpselt nii globaalsel kui ka regionaalsel tasandil.

2.1.2 Numbriliste meetodite piirangud

Vaatamata suurtele edusammudele ilmaennustuses on numbrilistel meetoditel ilmaennustuses mitu probleemi. Esimene probleem numbriliste ilmaennustusmudelitega on nende tööle hakkamise aeg (inglise k. spin-up time), mis tähistab perioodi, mille jooksul mudel stabiliseerub pärast käivitamist. Spin-up aeg on vajalik, et mudel jõuaks tasakaaluolekusse ning hakkaks realistlikult atmosfääri protsesse simuleerima [2]. Enamasti on operatiivsete mudelite puhul see aeg vähemalt 6 tundi, mille jooksul mudeli väljundid ei ole täielikult usaldusväärsed. Teine probleem on see, et suurte andmemahude töötlemine nõuab

tohutut arvutusvõimsust. Meteoroloogiliste andmete maht on äärmiselt suur, see seab kõrged nõudmised arvuti mälule ja protsessorile [3]. Kuna numbriliste ilmaennustuste arvutuslik kulu on tavapärasele arvutitele liiga suur, siis on tarvis operatiivseks kasutamiseks superarvuteid ja isegi neil on vaja rakendada paralleelarvutuse optimeerimismeetodeid [3]. Kolmandaks, numbriliste mudelite täpsus keskmise ja pika perspektiivi prognoosimisel on endiselt problemaatiline, eriti kiiresti muutuvate elementide, nagu 10m tuule ja õhurõhu osas [3].

2.1.3 ECMWF mudeli ülevaade

Euroopa Keskpika Ilmaennustuste Keskus (European Centre for Medium-Range Weather Forecasts, ECMWF) on üks maailma juhtivaid ilmaennustuskeskusi, mis pakub globaalseid numbrilisi ilmaennustusi. ECMWF mudel on tuntud oma suure täpsuse poolest [4].

ECMWF IFS-i (Integrated Forecasting System) ehk integreeritud prognoosimudelit (edaspidi kutsutud ECMWF mudeliks) iseloomustavad järgmised omadused: globaalne ulatus kogu Maa modelleerimiseks, kõrge horisontaalne lahutusvõime (ligikaudu 9 km) ja 137 vertikaalset kihti, mis ulatuvad kuni 0,01 hPa ehk umbes 80 km kõrgusele. ECMWF mudel on ansambelmudel (51 liiget), ehk sellega saab hinnata ka ennustuste määramatust ja luua tõenäosusprognoose. Mudel kasutab neljamõõtmelist variatsioonandmete assimilatsiooni (4D-Var), mis võimaldab kasutada suuri andmehulkasid nagu satelliitandmed. Lisaks on mudelisse integreeritud ookeani (NEMO), merejää (LIM), lainete (ECWAM) ja maismaa hüdroloogia (HTESSEL) mudelid kompleksete ennustuse tegemiseks

ECMWF IFS-i eeliseks on selle globaalne olemus ja väga suur arvutusvõimsus, mis võimaldab täpset modelleerimist ka keerukate füüsikaliste protsesside puhul. Piiranguks on aga piirkondlike iseärasuste arvestamine näiteks saarestiku ja rannikupiirkondades, kus kohalikud maastikutunnused ja mikrokliima võivad prognoositulemusi oluliselt mõjutada.

IFS mudelit uuendatakse regulaarselt ning selle operatiivprognoosid ulatuvad kuni 15 päevani (kõrglahutusega kuni 10 päeva). Ansambelprognoosi avaldatakse mitu korda nädalas ja hooajalisi prognoose mitu korda kuus.

2.1.4 MEPS mudeli ülevaade ja eripärad

MEPS (MetCoOp Ensemble Prediction System) on ansablennustussüsteem, mis põhineb HARMONIE-AROME mudelil – mittehüdrostaatilisel, kõrglahutusega regionaalsel mudelil, mis sobib hästi konvektiivsete protsesside modelleerimiseks [5]. METCoOp (Meteorological Co-operation on Operational Numerical Weather Prediction) on Põhja-Euroopa riikide (Norra, Rootsi, Soome, Eesti, Leedu) koostöö, mille eesmärgiks on ühise ilmaennustussüsteemi arendamine ja kasutamine [6].

MEPS mudeli põhiomadused on eriti kõrge horisontaalne lahutus (2.5 km), 65 vertikaalset kihti atmosfääris ja piiratud alaga mudeli struktuuri (Põhja-Euroopa). Mudel kasutab

ansamblennustust (30 liiget) ebakindluse hindamiseks, 3D-VAR andmete assimilatsiooni ja väljastab prognoose alates 2024. aastast ühe tunnise intervalliga.

MEPS mudeli eeliseks on kõrgem ruumiline lahutus võrreldes globaalsete mudelitega nagu ECMWF, mis võimaldab paremini arvestada kohaliku topograafia ja eripäradega. Mudel on optimeeritud Põhja-Euroopa ilmastikutingimuste modelleerimiseks, võttes arvesse piirkonna eripärasid. Erinevalt ECMWF mudelist, mis on üles ehitatud globaalsele skaalale, on MEPS disainitud regionaalsete ilmastikutingimuste täpsemaks modelleerimiseks (eriti konvektiivsete protsesside ja kohaliku topograafiaga seotud nähtuste osas).

2.2 Masinõpe ilmaennustuses: evolutsioon ja rakendused

2.2.1 Ajalooline areng ja praegune roll

Esimesed regressioonipõhised veakorrektsioonialgorimid meteoroloogias võeti kasutusele juba 1950 [7]. Sellest ajas saati on valdkonnas suuri arenguid olnud. Aja jooksul on valdkond järk-järgult omaks võtnud keerukamaid tehnikaid, seal hulgas juhuslikke metsi, gradiendivõimendamist ja viimasel ajal ka süvaõppe lähenemisviise [8]. Süvaõppe meetodid on viimastel aastatel teinud suuri edusamme, võimaldades arendada keerukaid masinõppemudeleid, mis põhinevad ainult algtingimustel: esmalt sademete ennustamiseks [9, 10, 11, 12] ja hiljem kogu atmosfääri erinevate parameetrite keskmise ajaulatusega ennustusteks [13, 14, 15, 16]. Erinevalt traditsioonilistest füüsikalistel võrranditel põhinevatel numbrilistest mudelitest suudavad masinõppe algoritmid tuvastada andmetes keerulisi mustreid ilma füüsikaseadusi otseselt modelleerimata [17]. See võimekus on muutnud masinõppe ilmaennustustes üha olulisemaks mitmes valdkonnas: numbriliste mudelite väljundite järeltöötlus [18], mõõtmisandmete lünkade täitmine, äärmuslike ilmastikunähtuste prognoosimine [19, 20] ning kiirete prognoosimudelite loomisel, mis vajavad vähem arvutusvõimsust [21].

2.2.2 Algoritmid, koolitusmeetodid ja väljakutsed

Meteoroloogilistes rakendustes on tõhusateks osutunud mitmed masinõppe meetodid, sealhulgas süvaõppe mudelid nagu konvolutsioonilised ja rekurrentsed närvivõrgud [22], gradiendi võimendamise algoritmid, näiteks XGBoost, LightGBM ja CatBoost [23] ning ka meetodid nagu Random Forest [24]. Käesolevas töös valiti nende hulgas süvaõppe perekonda kuuluv närvivõrgu meetod, selle mitmete eelise tõttu. Esiteks võimaldab see tõhusalt käsitleda muutujate vahelisi mittelineaarseid seoseid ning töödelda ja õppida suurest hulgast ajaloolistest ilmastikuandmetest. Lisaks pakub mudeli arhitektuur suurt paindlikkust, võimaldades kohandamist erinevate meteoroloogiliste nähtuste käsitlemiseks.

Kuigi paljud uuringud eelistavad libisevat treeningakent, kus testandmete periood muutub treenimise ajal, kasutavad operatiivsed järeltöötlusmudelid sagedamini staatilist treeningperioodi [25]. Uuringud on näidanud, et pikaajaliste treeningandmete kasutamine annab sageli paremaid tulemusi isegi juhul, kui aluseks olevas numbrilises mudelis või meteoro-

loogilistes tingimustes on toimunud muutusi [26]. Seetõttu valiti ka selles töös staatiline pikaajaline treeningperiood.

Süvaõppe meetodid, eriti konvolutsioonilised närvivõrgud, sobivad meteoroloogia jaoks erakordselt hästi, kuna neid saab rakendada otse aegruumilistele võrkudele, võimaldades oluliste tunnuste tuvastamist erinevatel skaaladel [8]. Siiski kaasnevad nendega ka mitmed väljakutsed. Nende meetodite rakendamine eeldab suuri andmekogumeid ja märkimisväärsed arvutusressursse. Samuti on tulemuste füüsikaline tõlgendamine keeruline [18], mistõttu on tarvis hoolikas valideerimine, et vältida ajalooliste mustritega liigsobitamist.

2.2.3 Korrektsioonimudelid ja arvutuslike väljakutsete lahendamine

Selles töös kasutatakse korrektsioonimudeli lähenemisviisi olemasolevate ilmaprognooside täiustamiseks, mitte asendusmodelite loomiseks. Selline strateegia on valitud, kuna globaalsed ja regionaalsed ilmaennustusmudelid ei arvesta piisavalt kohalike iseärasustega, näiteks rannikuala mõju või keeruka topograafiaga seotud nähtustega. Lisaks on masinõppe mudelid võimelised tuvastama ja korrektselt käsitlema süstemaatilisi vigu, mis esinevad traditsioonilistes prognoosimudelites [27].

Traditsioonilise numbrilise ilmaennustuse võimsuse ja täpsuse areng, mis järgib arvutustehnoloogia arengut, on lõppemas, kuna esiteks arvutikiiruse kasv sama hinna eest on aeglustumas (Moore'i seadus ja Dennardi skaleerimine) [28] ning teiseks numbrilised ilmaennustusprogrammid ei suuda olemasolevast tehnoloogiast rohkem jõudlust välja pigistada. See tähendab, et edasised täiustused vajavad suuremaid ja kallimaid masinaid, mille soetamis- ja käitamiskulud võivad muutuda operatiivsetele ilmakeeskustele taskukohatuks, kui prognoosimudelite nõuded kasvavad samas tempos kui varem [29, 21].

Masinõppe pakub alternatiivseid võimalusi või väljapääsu sellest probleemist. Tänapäeval kättesaadavad tohutud andmehulgad ja arvutusvõimsused on muutnud 20. sajandi masinõppe 21. sajandi süvaõppeks [29].

3. Metoodika

3.1 Andmete kogumine

Käesolevas töös kasutati nelja ilmajaama andmeid Tallinna lahe piirkonnas: Pirita, Tallinn-Harku, Rohuneeme ja Naissaare. Need jaamad valiti nende strateegilise asukoha põhjal, et saada terviklik ülevaade Tallinna lahe meteoroloogilistest tingimustest. Pirita jaam asub otse rannikul lahe kaguosas ning esindab linna lähedaste rannikualade tingimusi. Tallinn-Harku jaam paikneb sisemaal lahest edelasuunas ja pakub võrdlusandmeid mandrilisemate tingimuste kohta. Rohuneeme jaam asub Viimsi poolsaarel, lahe idas, esindades avatud rannikuala tingimusi ja Naissaare jaam paikneb Tallinna lahe keskel või põhjaosas, andes ainulaadse ülevaate puhtalt mereliste tingimuste kohta. Jaamade geograafiline jaotus võimaldab analüüsida erinevaid meteoroloogilisi mikrokliimasid, mis on iseloomulikud Läänemere rannikupiirkonnale. See on eriti oluline tuule parameetrite uurimiseks, kuna rannikulähedased alad on tuntud oma kiiresti muutuvate tuuleolude poolest, mis tulenevad maa-mere soojusvahetusest ja kohalikest topograafilistest mõjudest.

Ilmajaamade andmed koguti perioodilt 2014 kuni 2024 Eesti Keskkonnaagentuurilt [30]. Peale ilmajaamade andmete koguti ka ECMWF deterministliku mudeli ajaloolised ennustuste väärtused maale kõige lähemal tasandil aastatest 2014 kuni 2024 MARS arhiivist [31]. ECMWF arhiveeritud andmete allalaadimiseks protsess nõudis spetsiaalset luba ECMWF-lt. Andmed laeti alla tunnise intervalliga kuni 24-tunnise prognoosipikkuseni, et võimaldada analüüsi kogu lühiajalise ennustuste spektri ulatuses. Lisaks kasutati töös aastatel 2020 kuni 2024 kogutud MEPS ansamblelennustuse kontrollmudeli (MEPS_{control}) ajaloolisi ennustusi maapinna tasandil [32]. MEPS andmete kogumise periood on lühem kuna andmete formaat ja kättesaadavus vahetus aastal 2020. Andmete hankimine toimus MET Norway THREDDS serveri kaudu. MEPS_{control} andmed laeti samuti alla tunnise intervalliga kuni 24-tunnise prognoosipikkuseni.

Nii ECMWF kui ka MEPS_{control} mudeli puhul analüüsiti järgmisi parameetreid:

- Õhutemperatuur 2 meetri kõrgusel (°C)
- Tuule kiirus 10 meetri kõrgusel (m/s)
- Tuule suund 10 meetri kõrgusel (°).

Mõlema ilmaennustusmudeli andmete puhul määrati igale ilmajaamale lähim mudeli punkt, kasutades geograafiliste koordinaatide võrdlust ja arvutades geodeetilise vahemaa põhjal minimaalne kaugus. ECMWF mudeli ruumiline resolutsioon on ligikaudu 9 km, mis tähendab, et igale ilmajaamale võis vastata sama mudeli punkt mitme naaberjaamaga. MEPS_{control} mudeli kõrgem ruumiline resolutsioon (2,5 km) võimaldas täpsemat geograafilist vastavust jaamade ja mudeli punktide vahel.

3.2 Võrdlusmetoodika

Mudelite prognooside kvaliteedi hindamiseks kasutati erinevaid statistilisi näitajaid, mis võrdlesid mudeli ennustusi tegelike mõõtmistega ilmajaamades. Kasutatavad statistilised näitajad olid:

- **RMSD (Root Mean Square Difference)** ruutkeskmine erinevus, mis rõhutab suuremaid erinevusi ja annab hea ülevaate ennustuse täpsusest:

$$\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2}, \quad (3.1)$$

kus f_i on prognoositud ja o_i mõõdetud väärtus.

- **BIAS** keskmine viga näitab süstemaatilist kõrvalekallet tegelikest väärtustest, positiivne väärtus viitab mudeli ülehindamisele ja negatiivne alahindamisele:

$$\text{BIAS} = \frac{1}{n} \sum_{i=1}^n (f_i - o_i). \quad (3.2)$$

- **Tsirkulaarne erinevus** tuule suuna hindamisel tuleb arvestada, et nurga 0° ja 360° vahel pole erinevust. Seetõttu kasutati tsirkulaarset erinevuse valemit:

$$\delta_i = (f_i - o_i + 180) \bmod 360 - 180, \quad (3.3)$$

kus $\delta_i \in [-180, 180]$ ja positiivne väärtus näitab, et prognoos on päripäeva nihkes.

- **Normaliseeritud standardhälve** kontrollib, kas ennustatud andmete varieeruvus kattub mõõdetud andmete varieeruvusega.

$$\text{Standardhälve suhe} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - \bar{f})^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - \bar{o})^2}}, \quad (3.4)$$

kus \bar{f} on prognoositud väärtuste keskmine ja \bar{o} mõõdetud väärtuste keskmine.

- **Pearsoni korrelatsioonikordaja** mõõdab lineaarsust prognoositud ja mõõdetud väärtuste vahel. Väärtus 1 näitab tugevat positiivset lineaarset seost, 0 tähendab seose puudumist:

$$r = \frac{\sum_{i=1}^n (f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^n (f_i - \bar{f})^2} \sqrt{\sum_{i=1}^n (o_i - \bar{o})^2}}. \quad (3.5)$$

3.3 Korrektsioonimudel

3.3.1 Sissejuhatus

Numbriliste ilnamudelite süstemaatilised vead tekivad füüsikaliste protsesside lihtsustamisest, madalatest ruumilistest ja ajalistest resolutsioonidest ja keeruliste lokaalsete mõjude ebapiisavast kajastamisest [17]. Sellepärast on vajalik rakendada korrektsiooni-algoritme, mis suudavad õppida mudelite vigu ajalooliste andmete põhjal. Käesolevas töös rakendati süvaõppe meetodeid ECMWF ja MEPS_{control} mudelite täpsuse parandamiseks Tallinna lahe piirkonnas. Närvivõrkude valik tugineb nende võimele modelleerida keerulisi mittelineaarseid seoseid meteoroloogiliste parameetrite vahel ilma eelnevate eelduste vajaduseta. Erinevalt traditsioonilistest lineaarsetest korrektsioonimudelitest suudavad närvivõrgud automaatselt tuvastada peidetud mustreid andmetes, kohanduda hooajaliste variatsioonidega ning arvestada mitme parameetri koosmõjudega samaaegselt. See on eriti oluline rannikupiirkondades, kus kohalikud mõjud võivad oluliselt erineda traditsiooniliste mudelite prognoosidest. Analüüsisis käsitletakse kõiki mõõtmisjaamade andmete keskmisi tulemusi, et vältida liiga spetsiifiliseks minemist, ning et vältida üksikute jaamade mõõtmisvigade liialt andmetesse toomist.

3.3.2 Mudeli arhitektuur ja disain

Korrektsioonimudel kasutab süvaõppe meetodit, mis on optimiseeritud andma ilmaandmeid väikseima keskmise vea, parima normaliseeritud standardhälbega ja suurima korrelatsiooniga mõõtmisandmetega. Mudeli ehitamiseks kasutati Python programmeerimiskeeles mitmeid teke, sealhulgas TensorFlow/Keras närvivõrgu jaoks, Scikit-learn andmete eeltöötluks ja tulemuste hindamiseks, Pandas teeki andmete töötlemiseks ning Matplotlib ja Seaborn teke visualiseerimiseks. Selleks, et integreerida närvivõrk Scikit-learn raamistikku kasutati ka KerasRegressor wrapper-klassi.

Rakendati närvivõrgu mudel mitme peidetud kihiga, mille arhitektuur varieerus parameetri tüübi järgi. Temperatuuri jaoks kasutati kolme varjatud kihti (64, 32, 16 neuronit) keskmise kompleksuse jaoks. Tuule komponentide (u, v) jaoks kasutati keerulisemat struktuuri kolme suurema varjatud kihiga (96, 64, 32 neuronit), et saada paremini aru komplekssetest suhetest tuule mustrites.

3.3.3 Parameetrite valik ja kalibreerimine

Mudeli parameetrid valiti vaadates, milliseid väärtuseid kasutavad teised ilnamudelite parandamiseks [lam2023] ja nende katsetus ja testimis meetodil väiksema võimsusega masinale korrigeerimisel. Treenimise protsessis kasutati optimiseerijat õppimiskiirusega 0.001 koos muutuva partiide suurusega olenevalt andmete mahust. Varajane peatumine rakendati siis, kui 15 epohhi järjest ei toimunud paranemist, ning õppimiskiirus jagati kahega iga 7 epohhi järel ilma paranemiseta, miinimumkiirusega 10^{-5} .

Andmete jaotus vastas 80% treenimisandmetele, 10% valideerimisandmetele ja 10% testimisandmetele ECMWF mudeli jaoks, ehk neid ei jaotatud täpselt aastate kaudu ära, aga üldiselt võib öelda, et ECMWF mudeli treenimiseks kasutati 2014-2022 aasta andmeid, 2023 aasta andmeid valideerimiseks ja 2024 aasta andmeid testimiseks. MEPS_{control} mudeli puhul kasutati treenimiseks 2020-2022 andmeid, valideerimiseks 2023 andmeid ja testimiseks 2024 andmeid.

Andmete eeltöötuses eemaldati numbriliste ning kategooriliste väärtuste korral puuduvad väärtused. Peamiselt esines neid Eesti mõõtejaamade andmetes, kokku oli neid vähem kui 2%.

3.3.4 Hetkennustuste mudel

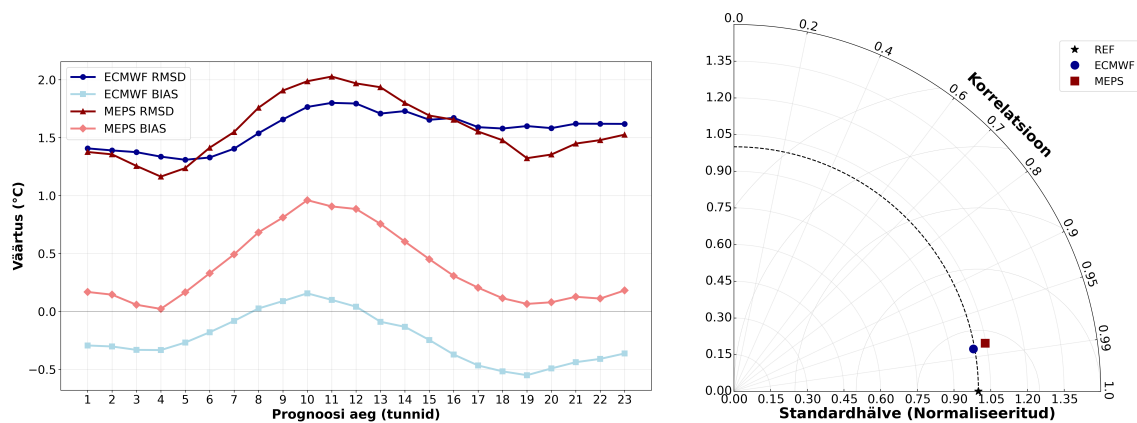
Lisaks tavalisele mudelile, mis ennustab järgmise 24 tunni väärtuseid treeniti ka hetkennustuste mudel, mis kasutab ka ilmajaamas reaalajas saadud andmeid ennustuste tegemisel. Tulemuste graafikud on sellel mudelil tehtud küll ööpäeva jaoks, aga päriselt saaks seda mudelit rakendada ainult üks tund ette, et saada järgneva tunni väga täpne ennustus. Seda saab ainult üks tund ette kasutada, kuna mudelis on tarvis eelnevate tundide ilmajaamade mõõtmistulemusi, ning neid ei ole tulevikust võtta päris ennustuse tegemisel ajal. Seda mudelit saaks iga tunni tagant uuesti jooksutada, et teada saada täpsemalt, mis ilm on lähiajal.

Hetkennustused oleksid tähtsad mitmes valdkonnas. Purjetamisvõistlustel, kus on võimalik arvuti või telefon merele kaasa võtta, aitaks järgneva tunni täpne tuuleinfo paremaid strateegilisi otsuseid teha ja annaks eelise võistlejate ees, kellel seda infot ei ole. Lennujaamades kaotavad lennufirmad suures koguses raha, kui nad peavad ilma või potentsiaalse tuleviku ilma tõttu reise tühistama [33]. Samuti oleksid täpsed lühiajalised ennustused väärtuslikud suuritel inimeste kogunemistel, nagu näiteks festivalid või tähtpäevad. Ehituses või põllumajanduses on tähtis mingeid etappe ainult kindlate ilmadega teha, näiteks betooni valamine või pestitsiidide taimedele panek, mistõttu oleks hetkennustused ka nendes valdkondades kasulikud.

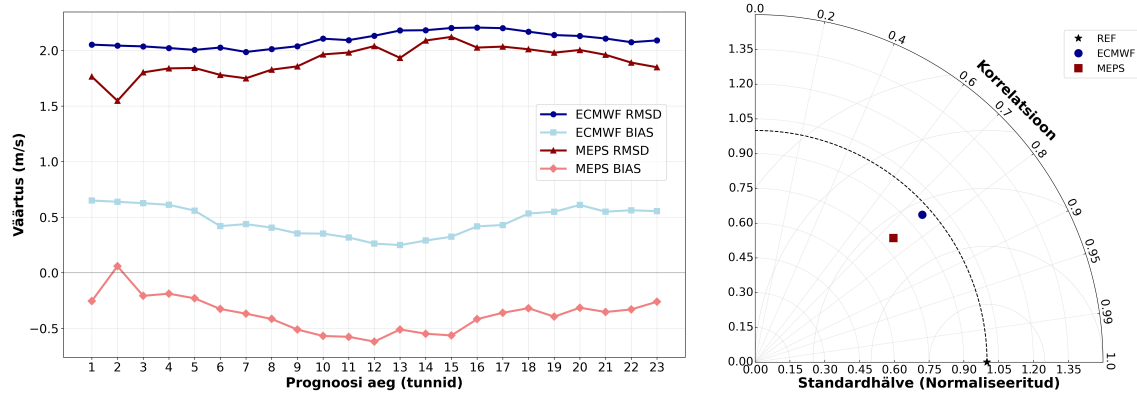
4. Tulemused

4.1 Esialgsete ilmaennustusmodelite võrdlus

Antud töös viidi läbi ECMWF ja MEPS_{control} mudelite võrdlus perioodil 2023 ja 2024. Selles uuriti mudelite täpsust temperatuuri, tuule suuna ja tuule kiiruse prognoosimises erinevate meetoditega. Joonisel 1 on esitatud ECMWF ja MEPS_{control} temperatuurienrustuste RMSD ja BIAS väärtused 23 tunnise prognoosipikkuse lõikes. Lisaks sellele on joonisel samade mudelite Taylor diagrammid, mis näitavad samuti RMSD-d, ning peale selle ka mudelite korreleeruvust mõõdetud andmetega. Taylori diagrammidel on veel näha normaliseeritud standardhälvet ehk seda, kui hästi langevad kokku mudelite ning mõõdetud andmete standardhälbed. Vasakpoolselt jooniselt on näha, et RMSD väärtused on MEPS_{control} mudelil paremad prognoosi alguses ja lõpus, aga tundide 6 kuni 16 vahel on ECMWF ruutkeskmine erinevus väiksem. MEPS_{control} mudeli temperatuuri kõrvalekalle on alguses parem, kui ECMWF mudeli oma, aja muutudes muutub halvemaks (hindab kuni kraadi võrra üle) ja prognoosiperioodi lõpuks on jälle parem. Mõlemad mudelid ülehindavad temperatuuri väärtuseid ennustuse keskmisel perioodil. ECMWF mudel alahindab temperatuuri nii prognoosi algul ja lõpul. Taylori diagrammilt on näha, et korrelatsioonid mõõdetud andmetega on mõlemal mudelil temperatuuri ennustamisel väga täpsed (99% lähedal). ECMWF mudeli standardhälve langeb mõõtmisandmetega peaaegu täpselt kokku ja MEPS_{control} mudeli tulemuste standardhälve on natukene suurem, kui ilmajaama omadel.

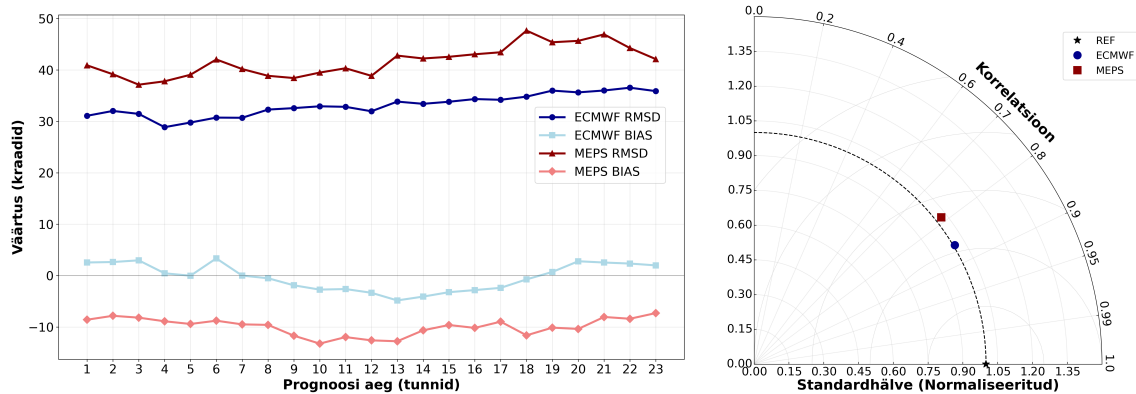


Joonis 1. ECMWF ja MEPS_{control} mudelite temperatuurienrustuste kvaliteet ilmajaamade keskmisega võrreldes Tallinna lähel (2023-2024 vaatlusandmed). Vasakul: MAE ja RMSD muutus 23 tunnise prognoosipikkuse ulatuses. Paremäl: Taylor diagramm mõlemale mudelile.



Joonis 2. ECMWF ja MEPS_{control} mudelite tuule kiiruse kvaliteet ilmajaamade keskmisega võrreldes Tallinna lähel (2023-2024 vaatlusandmed). Vasakul: MAE ja RMSD muutus 23 tunnise prognoosipikkuse lõikes. Paremalt: Taylor diagramm mõlemale mudelile.

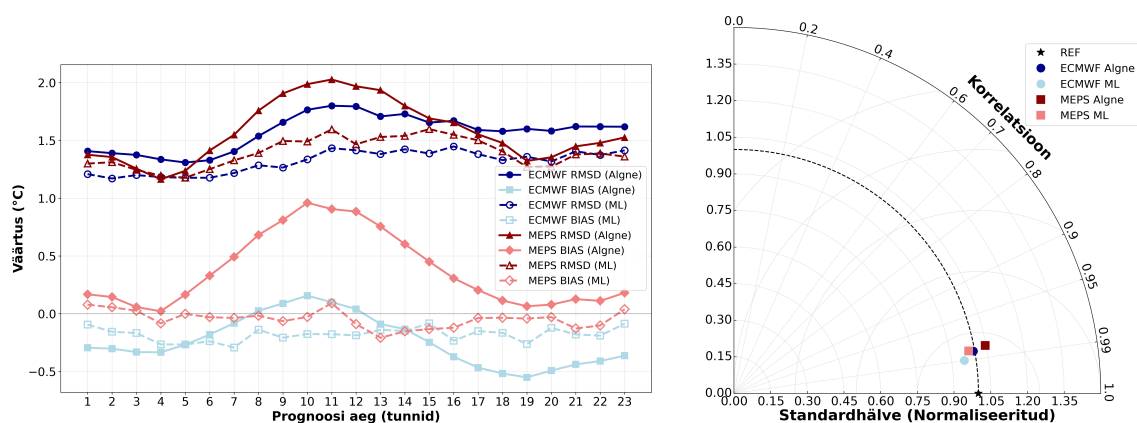
Lisaks temperatuurile koostati analoogsed graafikud ka mudelite tuule kiiruse võrdluse jaoks. Need on esitatud joonisel 2, ning neilt on näha, et MEPS_{control} näitab paremat jõudlust terve ennustuse vältel väiksema RMSD väärtusega. Algselt näitab MEPS_{control} minimaalset kõrvalekallet, samas kui ECMWF näitab 0.6 m/s suurt positiivset nihet. Prognoosiperioodi pikenedes see seos pöördub: MEPS_{control} kõrvalekalle liigub negatiivses suunas ja ulatub keskmise prognoosiperioodi jooksul umbes -0,6 m/s-ni, samas kui ECMWF nihe väheneb järk-järgult ja läheneb neutraalsetele väärtustele. Sarnaselt temperatuurile, mõlema mudeli kõrvalekalde väärtused liiguvad lõpuks sarnastele väärtustele, nagu need olid ennustusperioodi algul. Tuule kiiruse korrelatsioon on mõlemal mudelil kehvem võrreldes temperatuuri andmete korrelatsiooniga. MEPS_{control} normaliseeritud standardhälve on märgatavalt kehvem (umbes 0.77), kui ECMWF mudeli oma (umbes 0.96), ehk MEPS_{control} alahindab tuule kiiruse varieeruvust. MEPS_{control} paistab silma lühiajaliste (tunnid 1–5) ja pikemate (tunnid 17–23) ennustuste tegemisel, samas kui ECMWF näitab paremaid oskusi vahepealsel prognoosiperioodil. Mudelite kehvem korrelatsioon, kui temperatuuril näitab, et mõlemal mudelil on raskusi tuulekiiruse järjepideva prognoosi hoidmisega kogu perioodi jooksul. Vead võivad olla seotud päevaste tuule tsüklite ja Tallinna rannikukeskkonnale omase süsteemidega.



Joonis 3. ECMWF ja MEPS_{control} mudelite tuule suuna kvaliteet ilmajaamade keskmisega võrreldes Tallinna lähel (2023-2024 vaatlusandmed). Vasakul: MAE ja RMSD muutus 23 tunnise prognoosipikkuse lõikes. Paremalt: Taylor diagrammi mõlemale mudelile

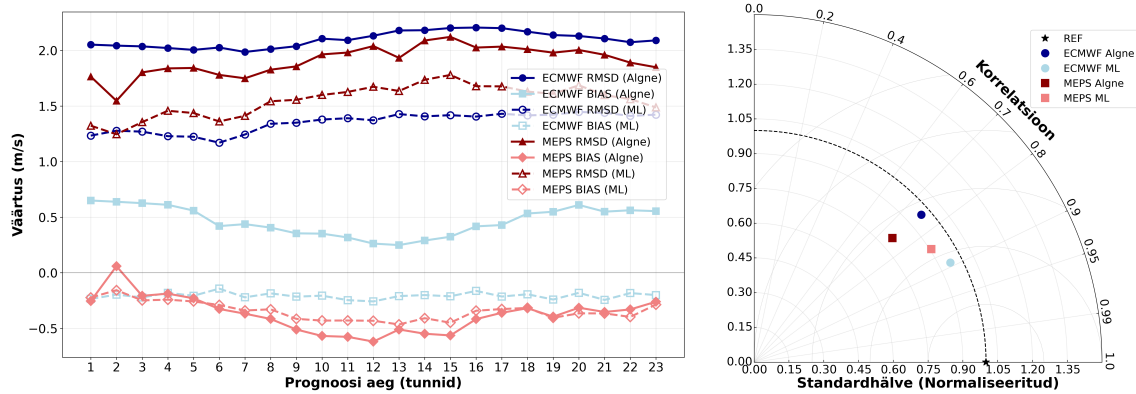
Tuule suuna graafikud on tehtud sarnaselt temperatuuri ja tuule kiiruse omadele, kus vasakul on RMSD ja BIAS väärtused võrreldes prognoosipikkusega, ning paremal on kahe mudeli võrdlus Taylor diagrammi abil. Need graafikud on joonisel 3 ja neist võib välja lugeda, et MEPS_{control} RMSD on ECMWF-i omast suurem terve prognoosiperioodi jooksul (38–48 kraadi), samas kui ECMWF-i ruutkeskmine erinevus on madalam (28–36 kraadi). Mõlema mudeli ruutkeskmised erinevused näivad olevat kasvavas tendentsis prognoosiperioodi vältel. ECMWF mudeli keskmine kõrvalekalle on tuule suuna jaoks nullilähedane. MEPS_{control} mudel prognoosib tuule suunda 9 kuni 12 kraadi vastupäeva nii prognoosiaja algul, keskel kui lõpul. Tuule suuna korrelatsiooninäitajad on mõlemal ilnamudelil temperatuuriprognosidega võrreldes kehvemad ja tuule kiiruse omadega sarnased. See tähendab, et mõlemal mudelil on raskusi tuule suuna järjepidevusega. ECMWF näitab kõrgemat korrelatsiooni kui MEPS_{control}. Mõlema mudeli standardhälbed on sarnased mõõdetud andmete omaga, ehk nad suudavad tabada loodusliku tuule suuna varieeruvuse. Halvemad statistilised seosed võrreldes temperatuuriga võivad viidata atmosfäärimustrite ja Tallinna rannikule omaste kohalike mõjude keerulisele sobitamisele. Halvemad tulemused võivad olla ka tuule suuna kiiresti muutuva omaduse tõttu, mida ainult tunnise resolutsiooniga ennustus ja vaatlusandmed ei suuda püüda.

4.2 Korrektsioonimudeli tulemused



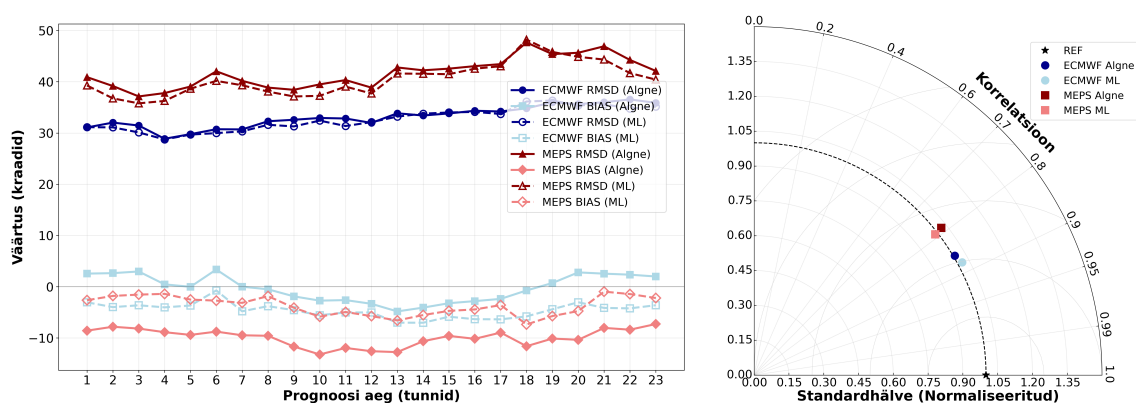
Joonis 4. Masinõppega parandatud temperatuuriennustuste kvaliteet ilmajaamade keskmisega võrreldes ECMWF ja MEPS_{control} mudelitel Tallinna lähel (2023-2024 vaatlusandmed). Vasakul: MAE ja RMSD muutus 23 tunnise prognoosipikkuse lõikes. Paremalt: Taylor diagramm kõikidele mudelile.

Jooniselt 4 on esitatud ECMWF ja MEPS_{control} algsete kui ka masinõppega korrigeeritud temperatuuriennustuste RMSD ja BIAS väärtused 23 tunnise prognoosipikkuse lõikes. Lisaks sellele on joonisel kõikide nelja mudeli Taylor diagrammid, mis näitavad samuti RMSD-d, aga ka parandatud mudelite korreleeruvust mõõdetud andmetega. Taylori diagrammidel on samuti näha normaliseeritud standardhälvet ehk seda, kui hästi langevad kokku nelja mudeli ning mõõdetud andmete standardhälbed. RMSD ja BIAS graafikult on näha, et mõlema mudeli ruutkeskmine viga temperatuuriprognosides väheneb masinõppe meetodite kasutamisel, ehk mudelite üldine täpsus suureneb. Masinõppe mudel on suutnud ECMWF temperatuuri kõrvalekallet stabiilsemaks saada, aga see eemaldas ka perioodid (7 ja 12 tunni juures), kus algmudel oli väga täpne. MEPS_{control} mudeli temperatuuri kõrvalekalde korral on masinõpe suutnud eemaldada kuni 1 °C ülehindamise 5 ja 17 tunni vahel. Korrigeeritud mudelitest on MEPS_{control} üldiselt väiksema kõrvalekaldega, aga ECMWF masinõppe mudel keskmiselt väiksema ruutkeskmise erinevusega. Taylori diagrammil asetsevad korrektsioonialgoritmidega saadud tulemused väga lähedal esialgsetele tulemustele ehk temperatuuri korral ei muutnud masinõppe märkimisväärselt varieeruvust ega lineaarset suhet alg- ja ennustatud väärtuste vahel.



Joonis 5. Masinõppega parandatud tuule kiiruse kvaliteet ilmajaamade keskmisega võrreldes ECMWF ja MEPS_{control} mudelitel Tallinna lähel (2023-2024 vaatlusandmed). Vasakul: MAE ja RMSD muutus 23 tunnise prognoosipikkuse lõikes. Paremal: Taylor diagramm kõikidele mudelile.

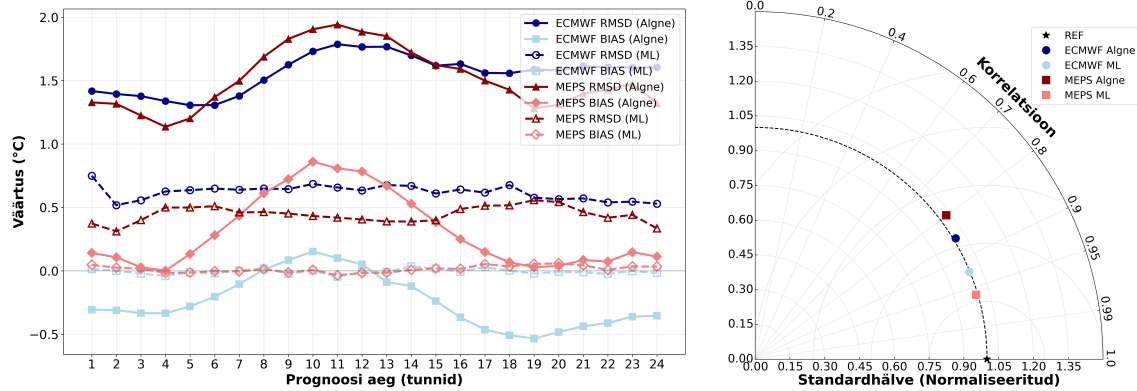
Sarnaselt eelnevale temperatuuri joonisele 4 on tehtud graafikud ka masinõppe mudelite tuule kiiruse prognoosi täpsuse hindamiseks, need asetsevad joonisel 5. Vasakpoolisel graafikul on näha, et masinõppe meetodid on ka tuule kiiruse korral suutnud ruutkeskmist erinevust vähendada mõlemal mudelil. Korrigeeritud ECMWF mudeli kõrvalekalle ning ruutkeskmise erinevus on mõlemad paremad MEPS_{control} mudeli omadest. Mõlema mudeli veakorrektoori algoritm stabiilselt alahindab tuule tugevust, nagu on näha negatiivsetest kõrvalekalletest. ECMWF parandatud mudelil on parim korrelatsioon kõikidest mudelitest ja ECMWF algmudelig sama normaliseeritud standardhälve. Korrigeerimismudel MEPS_{control}-le on suutnud parandada selle mõlemat Taylor diagrammi veanäitajat, ehk uue mudeliga saadud andmed korreleeruvad paremini algandmetega ning nende varieeruvus on ka sarnasem.



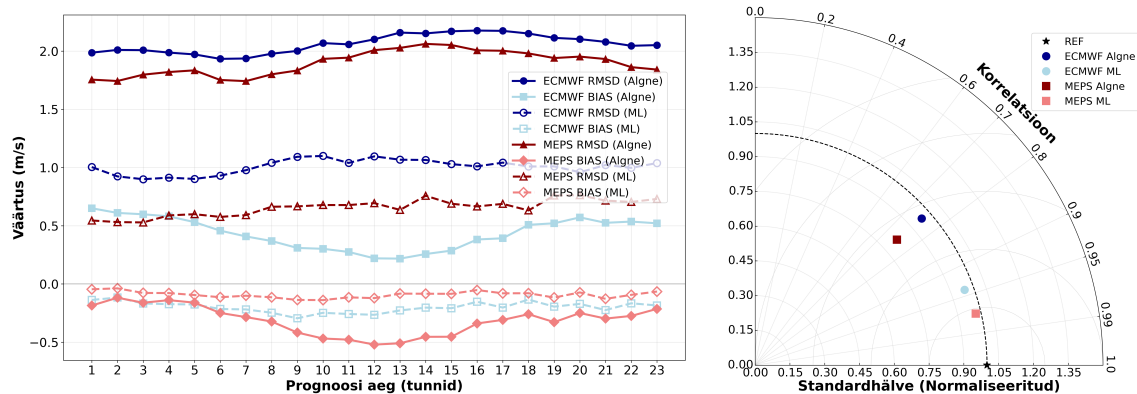
Joonis 6. Masinõppega parandatud tuule suuna kvaliteet ilmajaamade keskmisega võrreldes ECMWF ja MEPS_{control} mudelitel Tallinna lähel (2023-2024 vaatlusandmed). Vasakul: MAE ja RMSD muutus 23 tunnise prognoosipikkuse lõikes. Paremal: Taylor diagramm kõikidele mudelile

Lisaks temperatuurijoonisele 4 ja tuule kiiruse joonisele 5 on tehtud analoogsed graafikud ka masinõppe mudelite tuule suuna prognoosi täpsuse hindamiseks, need asetsevad joonisel 6. Sellel joonisel on masinõppega arenenud ainult MEPS_{control} kõrvalekalle. Mõlema mudeli ruutkeskmise erinevuse ei ole masinõppega nähtavalt paranenud ja ECMWF parandatud mudeli kõrvalekalle on mingitel ennustusperioodidel isegi suurenenud. On näha, et tuule suuna parandamine masinõppega on keerulisem kui tuule kiiruse parandamine.

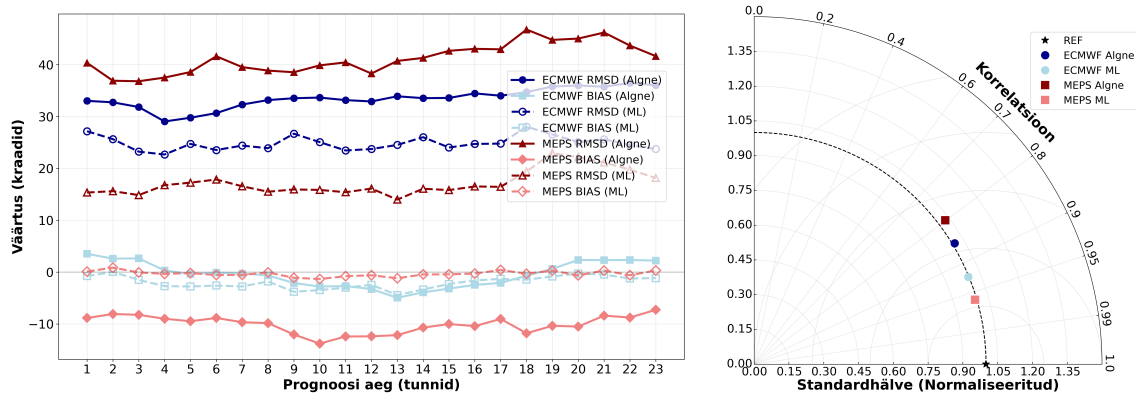
4.3 Hetkennustuse võrdlus



Joonis 7. Masinõppega parandatud hetkennustuse temperatuuri kvaliteet ilmajaamade keskmisega võrreldes ECMWF ja MEPS_{control} mudelitel Tallinna lähel (2023-2024 vaatlusandmed). Vasakul: MAE ja RMSD muutus 23 tunnise prognoosipikkuse lõikes. Paremalt: Taylor diagram kõikidele mudelile



Joonis 8. Masinõppega parandatud hetkennustuse tuule kiiruse kvaliteet ilmajaamade keskmisega võrreldes ECMWF ja MEPS_{control} mudelitel Tallinna lähel (2023-2024 vaatlusandmed). Vasakul: MAE ja RMSD muutus 23 tunnise prognoosipikkuse lõikes. Paremalt: Taylor diagram kõikidele mudelile



Joonis 9. Masinõppega parandatud hetkennustuse tuule suuna kvaliteet ilmajaamade keskmisega võrreldes ECMWF ja MEPS_{control} mudelitel Tallinna lähel (2023-2024 vaatlusandmed). Vasakul: MAE ja RMSD muutus 23 tunnise prognoosipikkuse lõikes. Paremalt: Taylor diagramm kõikidele mudelile

Masinõppega korrigeeritud mudelite võrdluse graafikud tehti ka hetkennustus mudelile jaoks sarnaselt eelnevatele. Graafikute loomisel kasutati mudelit, mis saab kasutada vahetult eelnevate tundide ilmajaamades mõõdetud andmeid järgnevate ennustuste tegemiseks. Joonistel on näha temperatuuri, tuule suuna ja tuule kiiruse parameetrite vigu ECMWF ja MEPS_{control} algmudelitel ja hetkennustuste parandatud mudelitel. Lisaks on näidatud ka hetkennustuste ja algmudelite asetsemine Taylor diagrammidel erinevate ilmaparameetrite korral. Hetkeilmaennustuse mudelit kasutada tahtes tuleb meeles pida, et see annab joonistel 7, 8 ja 9 näidatud täpsusega ennustuse ainult järgneva tunni jaoks. Neilt joonistelt on näha, et vastavale masinõppega parandatud ennustusmudelile eelnevate tundide mõõdetud ilmaandmete andmine suurendab täpsust märkimisväärselt. Hetkennustuse mudelid kasutavad viimaste mõõdetud andmete informatsiooni, et korrigeerida ennustuste vigu reaajas, mis annab olulise eelise võrreldes tavapärase parandatud mudelitega. See lähenemisviis vähendab kõrvalekalde peaaegu täielikult kõikide ilmaparameetrite korral. Selline lahendus suudab ka ruutkeskmist erinevust vähendada poolteist kuni kolm korda olenevalt mudelist ja ilmaparameetrist. Temperatuuri korral on parandus kõige märgatavam, kus mõlema mudeli RMSD väärtused langevad oluliselt. Tuule kiiruse prognoosides saavutatakse samuti märkimisväärne paranemine, ning on näha, et hetkennustuse mudelid suudavad kompenseerida nii üle- kui ka alahindamise vigu. Tuule suuna ennustamise täpsus paraneb samuti, kuigi vähemal määral kui teiste parameetrite puhul, mis on ootuspärane arvestades tuule suuna suurema varieeruvuse ja keerukama prognoosimise iseloomu. MEPS_{control} masinõppega korrigeeritud hetkennustuse mudel saavutab märkimisväärselt madalama ruutkeskmise erinevuse ECMWF korrigeeritud mudelist nii temperatuuri, tuule kiiruse kui ka tuule suuna korral, vaatamata algsetelt kehvemale ruutkeskmisele väärtusele kahes nendest kategooriatest. See näitab hetkennustuse lähenemisviisi potentsiaali madalamate resolutsioonidega mudelite täpsuse parandamisel. Taylori diagrammidelt on ka näha, et hetkennustuse mudelid langevad

mõõdetud andmetega märkimisväärselt paremini kokku kui algsed mudelid. Korrelatsiooni väärtused on kõrgemad kõikide parameetrite korral, eriti temperatuuri puhul, kus korrelatsioon läheneb peaaegu ideaalsele väärtusele. Normaliseeritud standardhälve on samuti paremini tasakaalus, mis näitab, et hetkennustuse mudelid varieeruvus kattub paremini mõõdetud andmete varieeruvusega. ECMWF hetkennustuse mudel säilitab oma algse hea korrelatsiooni, kuid vähendab oluliselt RMSD väärtusi, samal ajal kui MEPS_{control} hetkennustuse mudel parandab nii korrelatsiooni kui ka normaliseeritud standardhälvet, saavutades üldiselt tasakaalustavamama tulemuse Taylori diagrammi kõikides parameetrites.

4.4 Ennustuste kokkuvõte

Järgnevas tabelis on esitatakse üldistatud tulemused, mis on saadud kõikide 4 mõõtmisjaama ja kogu 23-tunnise prognoosiperioodi andmete koos analüüsimisel. Ühtse hindamissüsteemi loomiseks arvutati iga ilmaparameetri ja mudeli kombinatsiooni jaoks keskmine RMSD paranemine. Selline meetodika tagab, et tulemused ei ole kallutatud üksiku jaama või konkreetse prognoosipikkuse erandlikult heade või halbade tulemuste poolt. Selle asemel kajastavad need keskmised näitajad reaalselt operatiivset kasutust, kus ennustusmudel peab toimima efektiivselt kogu prognoosialal ja erinevatel aegadel.

Tabel 1. Keskmine RMSD Paranemise Protsent Masinõppega Parandatud Mudelitel Üle Algmudelite

Mudel	Näitaja	Tavaline närvivõrk	Hetkennustuse närvivõrk
ECMWF	Temperatuur	16.1%	60.2%
ECMWF	Tuule Suund	1.2%	26.2%
ECMWF	Tuule Kiirus	35.3%	51.2%
MEPS _{control}	Temperatuur	12.8%	71.0%
MEPS _{control}	Tuule Suund	2.7%	59.0%
MEPS _{control}	Tuule Kiirus	18.7%	66.0%

Tabelis esitatud tulemused näitavad selgeid mustreid mõlema lähenemisviisi efektiivsuses. Kõige märkimisväärsemat paranemist on saavutatud temperatuuri ennustamises, kus hetkennustuse närvivõrgud suudavad vähendada RMSD väärtusi 60-71% võrra. MEPS_{control} mudeli korral on temperatuuri paranemine suurem, kui ECMWF mudeli korral, mis viitab sellele, et madalamate resolutsioonidega mudelitel on suurem potentsiaal masinõppe meetodite rakendamisel. Tuule kiiruse prognoosides saavutatakse samuti olulised parandused, eriti hetkennustuse algoritmidega, kus RMSD paranemine ulatub 51-66% vahemikku. ECMWF mudeli puhul on juba tavapärane närvivõrk suuteline 35.3% paranemist saavutama, mis näitab selle mudeli head sobivust masinõppe meetoditega. Samas MEPS_{control} mudeli tuule kiiruse prognoosid paranevad hetkennustuse meetodil koguni 66.0%. Tuule suund osutub kõige keerukamaks parameetriks parandamisel, mis on ootuspärane arvestades selle tsirkulaarset iseloomu ja suuremat varieeruvust. Siiski on hetkennustuse lähenemisviis suuteline saavutama märkimisväärsed parandusi: ECMWF mudeli korral

26.2% ja MEPS_{control} mudeli puhul koguni 59.0%. Tavapäraseid närvivõrgud näitavad tuule suuna korral väiksemat efektiivsust, saavutades vaid 1-3% paremaid tulemusi, mis viitab sellele, et ainult statistiliste seoste kasutamine ilma reaajas mõõdetud andmeteta ei ole selle parameetri korral piisav.

5. Järeldused ja arutelu

Käesoleva töö ECMWF ja MEPS_{control} algsets mudelite võrdlus näitas, et temperatuuri ennustamisel saavutasid mõlemad mudelid väga kõrge korrelatsiooni mõõdetud andmetega. ECMWF mudel oli täpsem keskmise prognoosiperioodi jooksul, samas kui MEPS_{control} näitas paremat jõudlust prognoosi alguses ja lõpus. Tuule kiiruse prognoosides ületas MEPS_{control} mudel ECMWF mudeli jõudlust terve ennustusperioodi vältel, kuigi ECMWF näitas 0,6 m/s positiivset nihet alguses võrreldes MEPS_{control} mudeli -0,6 m/s nihkega keskmise perioodi jooksul. Tuule suuna ennustamisel oli ECMWF täpsem RMSD väärtustega 28-36 kraadi võrreldes MEPS_{control} mudeli 38-48 kraadiga, samuti näitas MEPS_{control} süstemaatilist 9-12 kraadi vastupäeva kõrvalekallet.

Masinõppe meetodite rakendamine tõi kaasa märkimisväärseid parandusi: temperatuuri RMSD paranes 12,8-16,1% tavapärase närvivõrkudega ja 60,2-71,0% hetkennustuse mudelitega. Tuule kiiruse prognoosid paranesid vastavalt 18,7-35,3% ja 51,2-66,0%. Tuule suuna ennustamine osutus kõige keerukamaks, saavutades 1,2-2,7% paranemise tavapärase meetoditega ja 26,2-59,0% hetkennustuse algoritmidega. MEPS_{control} mudel näitas üldiselt suuremat paranemispotentsiaali kõigis parameetrites masinõppe meetodeid kasutades.

Käesoleva uurimistöö põhjal võib teha järgmised neli peamist järeldust:

1. **Mudelite võrdlev analüüs Tallinna lahel:** Läbi viidi süstemaatiline võrdlus ECMWF globaalse mudeli ja MEPS_{control} piirkondliku mudeli vahel Tallinna lahe ilmatingimuste prognoosimise osas, kasutades nelja mõõtmisjaama andmeid 23-tunnise prognoosiperioodi vältel. Analüüs hõlmas kolme põhilist ilmaparameetrit: temperatuuri, tuule kiirust ja tuule suunda.
2. **Algmudelite jõudluse tulemused:** Võrdluse tulemused näitasid, et ECMWF mudel saavutas üldiselt parema jõudluse temperatuuri ja tuule suuna ennustamisel, samas kui MEPS_{control} mudel oli täpsem tuule kiiruse prognoosides lühiajaliselt. Mõlemad mudelid näitasid head korrelatsiooni ja väikeseid vigu temperatuuri parameetri osas, kuid tuule parameetrite ennustamine osutus keerukamaks.
3. **Veakorrektsiooni algoritmide arendamine:** Töö käigus arendati välja kaks tüüpi masinõppel põhinevaid veakorrektsiooni algoritmi: tavapäraseid närvivõrgud ja hetkennustuse mudelid. Hetkennustuse lähenemisviis, mis kasutab reaajas mõõdetud andmeid, osutus märkimisväärselt efektiivsemaks, võimaldades saavutada kuni kolm korda paremat ruutkeskmist erinevust võrreldes algmudelitega.
4. **Parandatud mudelite üldtulemused:** Masinõppega korrigeeritud mudelid saavutasid märkimisväärseid parandusi kõikides ilmaparameetrites. Hetkennustuse algoritmid näitasid eriti häid tulemusi: temperatuuri ennustamise täpsus paranes

60-71%, tuule kiiruse ennustamine 51-66% ja tuule suuna ennustamine 26-59% võr-
ra. MEPS_{control} mudel, vaatamata algsetele kehvematele tulemustele, suutis pärast
masinõppe rakendamist üldiselt ületada ECMWF mudeli jõudlust kõikides parameet-
rites.

Käesoleva töö tulemused on kooskõlas rahvusvaheliste uuringutega, mis kinnitavad ma-
sinõppe meetodite efektiivsust ilmaennustuste parandamisel [17]. Saavutatud 12-16%
paranemine temperatuuri ennustamises ja 19-35% paranemine tuule kiiruse prognoosides
on sarnased tulemustega, mida on kirjeldatud post-processing meetodite uuringutes [18,
27]. Tuule suuna ennustamise keerukus, kus paranemine oli vaid 1-3%, on kooskõlas
kirjandusega, mis rõhutab tsüklilise parameetri eripärasid masinõppe rakendamisel.

Tulemused näitavad, et madalamate resolutsioonidega mudelitel (MEPS_{control}) võib olla
suurem potentsiaal masinõppe meetodite rakendamisel, mis on oluline leid operatiivsete
ilmateenuste arendamisel. Töö praktiline väärtus seisneb selle rakendatavuses Eesti ranni-
kupiirkondades, eriti purjetamise ja merenduse valdkondades, kus täpsed tuule prognoosid
on kriitilise tähtsusega. Väljatöötatud algoritmide ülekandmine teistesse rannikupiirkonda-
desse nõuaks täiendavat uurimist ja kohandamist kohalike ilmaeärasustega.

6. Kokkuvõte

Käesoleva uurimistö eesmärgiks oli hinnata ECMWF ja MEPS_{control} ilmapudelite kasutatavust Tallinna lahe piirkonnas ning töötada välja meetodika nende mudelite tulemuste täpsustamiseks. Kõik püstitatud uurimisülesanded said täidetud. ECMWF ja MEPS_{control} mudelite võrdlev analüüs näitas, et mõlemal mudelil on oma tugevused ja nõrkused Tallinna lahe piirkonna ilmastiku modelleerimisel. Mudelite süstemaatiliste kõrvalekallete analüüs tõi välja selged mustrid. ECMWF mudel kaldus temperatuuri alahindama ja tuule suunda päripäeva nihutama, samas kui MEPS_{control} mudel ülehindas temperatuuri ja nihutas tuule suunda vastupäeva. Masinõppel põhinev korrektsiooniviis ECMWF ja MEPS_{control} mudelite tulemuste täpsustamiseks osutus efektiivseks, parandades prognoose iga näitaja korral. Lähiajaline närvivõrguga parandatud ennustusmudel koos varasemate mõõdetud andmetega suutis ilmaennustuse täpsust märkimisväärselt parandada.

Töö vihjab potentsiaalselt kasulikele edasistele uurimissuunadele, nagu keerukamate mudeliarhitektuuride katsetamine või füüsikateadlike närvivõrkude kasutuselevõtt. Tulevased uuringud saaksid mudeli sisendite täiustamiseks kaasata ka satelliidi- ja ookeaniandmeid või laiendada piirkonda teistele Läänemere piirkondadele. Need arengud parandaksid veelgi masinõppega täiustatud ilmaennustussüsteemide täpsust ja rakendatavust.

Kasutatud kirjandus

- [1] Thomas Tomkins Warner. *Numerical Weather and Climate Prediction*. Cambridge University Press, 2011. ISBN: 9780521513890. DOI: 10.1017/CB09780511763243.
- [2] Malte Müller *et al.* "AROME-MetCoOp: A Nordic Convective-Scale Operational Weather Prediction Model". *Weather and Forecasting* 32.2 (2017), lk. 609–627. DOI: 10.1175/WAF-D-16-0099.1.
- [3] Yue Wang *et al.* "An integrated framework for spatio-temporal-spectral fusion and migration learning for global sea surface wind speed retrieval". *Computers Geosciences* 150 (2021), lk. 104726. DOI: 10.1016/j.cageo.2021.104726.
- [4] European Centre for Medium-Range Weather Forecasts. *Integrated Forecasting System (IFS) Documentation*. 2023. URL: <https://www.ecmwf.int/en/publications/ifs-documentation>.
- [5] Lene Østvand *et al.* *Verification of Operational Weather Prediction Models: December 2023 to February 2024*. METinfo 31/2024. Norwegian Meteorological Institute, mai 2024. URL: https://www.met.no/publikasjoner/met-info/met-info-2024/_/attachment/inline/00476eb6-c021-4601-beb5-87e29c8f96f9:8de05906dd220c52fa401e7c2f16491b569a6907/MET-info-31-2024.pdf.
- [6] METCoOp. *MetCoOp Ensemble Prediction System (MEPS)*. Accessed May 2025. 2023. URL: <https://www.met.no/en/projects/metcoop>.
- [7] Thomas F. Malone. "Application of statistical methods in weather prediction". *Proceedings of the National Academy of Sciences* 41.11 (1955), lk. 806–815. DOI: 10.1073/pnas.41.11.806.
- [8] Amy McGovern *et al.* "Using Artificial Intelligence to Improve Real-Time Decision-Making for High-Impact Weather". *Bulletin of the American Meteorological Society* 100.11 (2019), lk. 2073–2090. DOI: 10.1175/BAMS-D-18-0119.1.
- [9] Xingjian Shi *et al.* "Convolutional LSTM network: A machine learning approach for precipitation nowcasting". *Advances in Neural Information Processing Systems* 28 (2015), lk. 802–810.
- [10] Suman Ravuri *et al.* "Skillful precipitation nowcasting using deep generative models of radar". *Nature* 597.7878 (2021), lk. 672–677. DOI: 10.1038/s41586-021-03854-z.
- [11] Jussi Leinonen, Daniele Nerini ja Alexis Berne. "Nowcasting-NET: a physics-aware deep learning model for weather forecasting". *Journal of Advances in Modeling Earth Systems* 15.1 (2023), e2022MS003239. DOI: 10.1029/2022MS003239.
- [12] Xinglong Zhang *et al.* "Skillful nowcasting of extreme precipitation with NowcastNet". *Nature* 619 (2023), lk. 527–532. DOI: 10.1038/s41586-023-06184-4.

- [13] Ryan Keisler. “Forecasting Global Weather with Graph Neural Networks”. *arXiv preprint arXiv:2202.07575* (2022).
- [14] Jaideep Pathak *et al.* “FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators”. *arXiv preprint arXiv:2202.11214* (2022).
- [15] Kaifeng Bi *et al.* “Accurate medium-range global weather forecasting with 3D neural networks”. *Nature* 619 (2023), lk. 533–539. DOI: 10.1038/s41586-023-06185-3.
- [16] Remi Lam *et al.* “GraphCast: Learning skillful medium-range global weather forecasting”. *arXiv preprint arXiv:2212.12794* (2022).
- [17] Peter Bauer, Alan Thorpe ja Gilbert Brunet. “The quiet revolution of numerical weather prediction”. *Nature* 525.7567 (2015), lk. 47–55. DOI: 10.1038/nature14956.
- [18] Benedikt Schulz ja Sebastian Lerch. “Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison”. *Monthly Weather Review* 150.1 (2022), lk. 235–257. DOI: 10.1175/MWR-D-21-0150.1.
- [19] David John Gagne *et al.* “Interpretable Deep Learning for Spatial Analysis of Severe Hailstorms”. *Monthly Weather Review* 147.8 (2019), lk. 2827–2845. DOI: 10.1175/MWR-D-18-0316.1.
- [20] Yunjie Liu *et al.* “Application of deep convolutional neural networks for detecting extreme weather in climate datasets”. *arXiv preprint arXiv:1605.01156* (2016).
- [21] Peter Bauer *et al.* “Digital revolution of Earth-system science”. *Nature Computational Science* 3.8 (2023), lk. 656–669. DOI: 10.1038/s43588-023-00443-0.
- [22] Yann LeCun, Yoshua Bengio ja Geoffrey Hinton. “Deep learning”. *Nature* 521.7553 (2015), lk. 436–444. DOI: 10.1038/nature14539.
- [23] Yan Ju *et al.* “A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting”. *IEEE Access* 7 (2019), lk. 28309–28318. DOI: 10.1109/ACCESS.2019.2901920.
- [24] David John Gagne, Amy McGovern ja Ming Xue. “Storm-Based Probabilistic Hail Forecasting with Machine Learning Applied to Convection-Allowing Ensembles”. *Weather and Forecasting* 32.5 (2017), lk. 1819–1840. DOI: 10.1175/WAF-D-17-0010.1.
- [25] Richard Hess. “Statistical postprocessing of ensemble forecasts for severe weather at Deutscher Wetterdienst”. *Nonlinear Processes in Geophysics* 27 (2020), lk. 473–487. DOI: 10.5194/npg-27-473-2020.
- [26] Moritz N. Lang *et al.* “Distributed training of neural network models: a comparison of strategies for robust out-of-sample predictions”. *Monthly Weather Review* 148.1 (2020), lk. 371–388. DOI: 10.1175/MWR-D-19-0196.1.

- [27] John B. Bremnes. “Ensemble postprocessing using quantile function regression based on neural networks and Bernstein polynomials”. *Monthly Weather Review* 148.1 (2020), lk. 403–414. DOI: 10.1175/MWR-D-19-0227.1.
- [28] John Shalf. “The future of computing beyond Moore’s Law”. *Philosophical Transactions of the Royal Society A* 378.2166 (2020). DOI: 10.1098/rsta.2019.0061.
- [29] Sergey Frolov *et al.* “Towards a new paradigm for numerical weather prediction”. *Science in the Digital Era* 1 (2024), lk. 1–12. DOI: 10.1016/j.scidig.2023.09.003.
- [30] Eesti Keskkonnaagentuur. *Keskkonnaportaal – Avaandmed*. <https://keskkonnaportaal.ee/et/avaandmed>. Meteorological data (wind speed, direction, temperature) for the period 2014–2024 from stations in the Tallinn Bay area (Pirita, Tallinn-Harku, Rohuneeme, Naissaar). Accessed on 2025-05-22. 2024.
- [31] European Centre for Medium-Range Weather Forecasts (ECMWF). *ECMWF MARS Archive – Deterministic Forecast Data*. <https://www.ecmwf.int/en/forecasts/dataset/operational-archive>. Historical deterministic forecasts (2014–2024) at surface-nearest level retrieved via MARS archive with access provided through collaboration with Estonian national partners. 2024.
- [32] MET Norway. *MEPS Ensemble Forecasts – Control Member Data via THREDDS Server*. <https://thredds.met.no/>. Historical MEPS_{control} forecasts (2020–2024) at surface level accessed through the Norwegian Meteorological Institute’s THREDDS data server. 2024.
- [33] Andrew Cook, Graham Tanner ja Stephen Anderson. “European airline delay cost reference values”. *Performance Review Unit, Eurocontrol* (2004). Updated in 2015 report estimating costs between 25,000–35,000 EUR per hour for weather-related delays.

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Oskar Villem Tiidemann

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose “Operatiivsete ilnamudelite analüüs ja võrdlus Tallinna lahel”, mille juhendaja on Arne Männik
 - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

23.05.2025

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