



Vesihuoltolaitoksen koneope- tettu omaisuudenhallintamalli

Käyttöomaisuusinvestointien ja taloussuunnittelun auto-
matisointi

Tomi Teuho

OPINNÄYTETYÖ
Toukokuu 2026

Dataosaamisen ja tekoälyn ylempi tutkinto-ohjelma

TIIVISTELMÄ

Tampereen ammattikorkeakoulu
Dataosaamisen ja tekoälyn ylempi tutkinto-ohjelma

TEUHO, TOMI:
Vesihuoltolaitoksen koneopetettu omaisuudenhallintamalli
Käyttöomaisuusinvestointien ja taloussuunnittelun automatisointi

Opinnäytetyö 125 sivua, joista liitteitä 12 sivua
Toukokuu 2026

Vesihuollon omaisuudenhallinta on keskeinen osa julkisen infrastruktuurin kestävästä kehitystä. Vesihuoltolain uudistus astui voimaan 1.1.2026 ja velvoitti kaikki vesihuoltolaitokset laatimaan systemaattisen omaisuudenhallintasuunnitelman. Suunnitteluprosessi on monille laitoksille vaativa, sillä se edellyttää erityisosaamista, laajojen aineistojen käsittelyä sekä resursseja, joita erityisesti pienemmillä laitoksilla on rajallisesti käytettävissä.

Tämän opinnäytetyön tavoitteena oli kehittää koneoppimiseen perustuva malli, jolla omaisuudenhallintasuunnitelman laatimista ja päivittämistä voidaan automatisoida ja tehostaa. Tutkimus toteutettiin suunnittelutieteellisenä konstruktio tutkimuksena, jossa matemaattinen mallinnus, syväoppiminen ja taloudellinen analyysi yhdistettiin yhtenäiseksi Python-pohjaiseksi ohjelmistokokonaisuudeksi.

Kehitetty artefakti perustui modulaariseen kolmikerroksiseen arkkitehtuuriin, joka sisälsi vesihuoltolaitoksen tilinpäätös- ja investointisuunnittelutietojen keruun sekä niiden yhdistämisen, koneoppimiseen perustuvan ennustamisen ja automatisoidun raportoinnin Word-tiedostoksi. Vesihuoltolaitoksen taloussuunnitteluun liittyvä inflaatioennuste toteutettiin tulkittavalla N-BEATS-neuroverkolla ja kustannusennuste yleisellä N-BEATS-neuroverkolla. Taloudellista kestävyyttä arvioitiin kassavirtaperusteisen liikevaihtotavoitteen sekä DuPont-analyysiin ja Altmanin Z"-malliin perustuvan riskinarvioinnin avulla.

Tulokset osoittivat, että vesihuoltolain edellyttämät omaisuudenhallintasuunnitelman keskeiset elementit voidaan tuottaa suurelta osin automaattisesti olemassa olevasta tilinpäätös- ja investointiaineistosta. Taloudellinen kannattavuus muodostaa omaisuudenhallinnan perustan, koska kannattava liiketoiminta mahdollistaa saneerausten toteuttamisen ja uusinvestointien rahoittamisen pitkällä aikavälillä. Kehitetty ratkaisu vapauttaa asiantuntijatyöaikaa rutiinilaskennasta analyytiseen tarkasteluun ja tukee läpinäkyvää päätöksentekoa. Työ tarjoaa konkreettisen ratkaisumallin vesihuoltolaitosten taloudenhallinnan ja strategisen investointisuunnittelun tueksi sekä vesihuoltosektorin digitaalisen murroksen edistämiseksi.

Asiasanat: vesihuolto, omaisuudenhallinta, koneoppiminen, neuroverkot

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Master's Degree Programme in Data Expertise and Artificial Intelligence

TEUHO, TOMI:
Machine-Learned Asset Management Model for a Water Utility
Automating Capital Investment and Financial Planning

Master's thesis 125 pages, appendices 12 pages
May 2026

Asset management in water services is a key component of the sustainable development of public infrastructure. The reform of the Water Services Act entered into force on 1 January 2026 and requires all water utilities to prepare a systematic asset management plan. For many utilities, the planning process is demanding, as it requires specialized expertise, the processing of extensive datasets, and resources that are particularly limited in smaller utilities.

The objective of this thesis was to develop a machine learning-based model for automating and streamlining the preparation and updating of asset management plans. The study was conducted as design science research, in which mathematical modeling, deep learning, and financial analysis were integrated into a unified Python-based software system.

The developed artifact was based on a modular three-layer architecture that included the collection and integration of the water utility's financial statement and investment planning data, machine learning-based forecasting, and automated reporting into a Word document. The inflation forecast for financial planning was implemented using an interpretable N-BEATS neural network, while cost forecasting was carried out using a generic N-BEATS neural network. Financial sustainability was assessed through a cash flow-based revenue target and a risk assessment based on DuPont analysis and the Altman Z" model.

The results showed that the key elements of the asset management plan required by the Water Services Act can be produced largely automatically from existing financial statement and capital investment data. Financial profitability forms the foundation of asset management, as a profitable business enables the implementation of renovations and the financing of new investments over the long term. The developed solution shifts expert time from routine calculations to analytical work and supports transparent decision-making. The study provides a concrete solution model to support financial management and strategic investment planning in water utilities, as well as to advance the digital transformation of the water sector.

Key words: water supply, asset management, machine learning, neural networks

TEKOÄLYN KÄYTTÖ OPINNÄYTTEESSÄ

Opinnäytteessäni on käytetty tekoälysovelluksia:

- Ei
 Kyllä

Ilmoitukseni mukaan olen käyttänyt opinnäytteessäni opinnäytetyöprosessin aikana seuraavia tekoälysovelluksia:

Tekoälysovellusten nimet ja versiot: Claude (Anthropic), versiot Claude Sonnet 4.6 ja Claude Opus 4.7

Käyttötarkoitus: Tekoälyä on hyödynnetty Python-ohjelmakoodin laatimisessa. Tekoälyavusteisessa koodauksessa tekijä on kuvannut tekoälylle luonnollisella kielellä halutun toiminnallisuuden, jonka perusteella tekoäly on tuottanut sitä vastaavan koodiehdotuksen. Tekijä on toiminut ohjelmistokehittäjän roolissa ja ohjannut prosessia vaiheittain täsmentämällä vaatimuksia, arvioimalla tuotettua koodia ja antamalla tekoälylle korjaavaa palautetta.

Osiot, joissa tekoälyä on käytetty: Tekoälyä on hyödynnetty ohjelmakoodin tuottamisessa, vaihtoehtoisten ratkaisujen vertailussa, virheidenkorjauksessa sekä koneoppimiskirjastojen toiminnan selvittämisessä ja matemaattisten kaavojen tarkistamisessa. Ohjelmistokokonaisuuden varsinainen suunnittelu, testaus ja toiminnallinen validointi on tehty itsenäisesti. Koodin lopullinen sisältö sekä kaavojen tulkinta ja niiden kytkentä teoreettiseen viitekehukseen perustuvat tekijän omaan asiantuntemukseen ja käytettyihin lähteisiin.

Olen tietoinen siitä, että olen täysin vastuussa koko opinnäytteeni sisällöstä, mukaan lukien osat, joissa on hyödynnetty tekoälyä, ja hyväksyn vastuun mahdollisista eettisten ohjeiden rikkomuksista.

SISÄLLYS

1	JOHDANTO	10
2	TUTKIMUKSEN TIETOPERUSTA.....	13
	2.1 Tutkimuksen tavoitteet	13
	2.2 Tutkimusmenetelmät.....	14
	2.2.1 Suunnittelutieteellinen konstruktio tutkimus	14
	2.2.2 Tutkimusmenetelmien normatiivinen konteksti	15
	2.2.3 Matemaattinen mallinnus ja simulointi	16
	2.3 Aineisto ja tietolähteet	17
	2.4 Tutkimuskysymykset	19
	2.5 Työn rajaukset.....	19
	2.6 Keskeiset käsitteet	20
3	OMAISUUDENHALLINNAN TALOUSHALLINNUS.....	22
	3.1 Aikasarja-analyysi	22
	3.1.1 Stokastiset prosessit ja aikasarjat.....	22
	3.1.2 Aikasarjan dekomponointi	23
	3.1.3 Inflaatio	25
	3.2 Vesihuoltolaitoksen investointi- ja taloussuunnittelu.....	27
	3.2.1 Vesihuoltolaitoksen omaisuudenhallinta.....	27
	3.2.2 Vesihuoltolaitoksen investointisuunnittelu	29
	3.2.3 Vesihuoltolaitoksen taloussuunnittelu	31
	3.3 DuPont-analyysi.....	33
	3.3.1 ROTA-hinnoittelumalli taloussuunnittelussa	34
	3.3.2 Liikevaihtotavoitteen matemaattinen malli	36
	3.3.3 Kokonaisvarallisuuden keskiarvoanalyysi.....	37
	3.3.4 Kustannusten ja liikevaihdon inflaatiodynamiikka	38
	3.3.5 Liikevoittotavoite omaisuuden arvon säilyttämiseksi.....	40
	3.3.6 Soveltaminen investointisuunnittelussa	43
4	KONEOPPIMISEN PERIAATTEET	45
	4.1 Python-ohjelmointikieli	45
	4.2 Koneoppiminen	46
	4.3 Syväoppiminen.....	49
	4.4 Syväoppimisen kehitysympäristö	52
	4.4.1 PyTorch-syväoppimiskehys	52
	4.4.2 PyTorch Lightning ja PyTorch Forecasting	53
	4.4.3 N-BEATS-neuroverkkoarkkitehtuuri.....	55
5	KONEOPETETTU OMAISUUDENHALLINTAMALLI	60

5.1	Artefaktin kehittäminen.....	60
5.2	Arkkitehtuuri	61
5.2.1	Tiedonkeruu ja yhdistäminen.....	62
5.2.2	Koneoppimismoduulit	63
5.2.3	Visualisoinnin merkitys	64
5.3	Inflaatioanalyysi ja pitkän aikavälin inflaatioennuste	64
5.3.1	Trendin estimointi paloittaisella lineaarisella regressiolla ...	67
5.3.2	CEEMDAN-algoritmi.....	69
5.3.3	Lookback-ikkunan valinta	72
5.3.4	Mallin koulutus ja validointi	74
5.3.5	Ennusteet	79
5.4	Pitkän aikavälin investointisuunnitelma	83
5.4.1	Investointien rahoitus.....	85
5.4.2	Investointien poistosuunnitelma.....	87
5.5	Koneopetettu taloussuunnitelma	89
5.5.1	Operatiivisten kustannusten ennustaminen	89
5.5.2	Pitkän aikavälin liikevaihtotavoitteen asettaminen	91
5.5.3	Vesihuollon maksujen automaattinen hinnoittelu	95
5.5.4	Riskinarvio vesihuoltolaitoksen talouden tilasta	100
5.6	Automatisoitu omaisuudenhallintasuunnitelman dokumentointi .	104
6	JOHTOPÄÄTELMÄT JA POHDINTA.....	106
	LÄHTEET.....	110
	LIITTEET	114
	Liite 1. Euribor 12 kk -viitekoron haku ECB Data Portal -palvelusta .	114
	Liite 2. Hilbert IMF -signaalianalyysi.....	115
	Liite 3. Inflaatiomallin koulutuksen diagnostiikka ja metriikat	116
	Liite 4. Inflaatiomallin validointijakson ja testikauden ennusteet	118
	Liite 5. Euribor 12 kk -viitekoron pitkän aikavälin estimointiennuste .	119
	Liite 6. Kustannusennustemallin diagnostiikka.....	120
	Liite 7. Riskinarvio – Case-vesihuoltolaitos	122

LYHENTEET JA TERMIT

aikasarja	ajan funktiona havaittujen arvojen joukko
aktivointifunktio	neuronin epälineaarinen muunnos, joka määrää, miten painotettu syötteiden summa välitetään eteenpäin
API	rajapinta sovellusten väliseen tietojen vaihtoon
artefakti	kehitetty ratkaisu kuten malli, menetelmä tai järjestelmä
backcast	mallin tuottama datarakenne syöteikkunasta
bootstrap	otantamenetelmä, jossa havaintojoukosta poimitaan toistuvasti satunnaisia näytteitä
CEEMDAN	signaalin hajotelma-algoritmi
data	jäsennelty tai jäsentymätön digitaalinen informaatio
deflaatio	rahan arvon nousu
dekomponointi	kokonaisuuden jakaminen pienempiin osiin
deterministinen	täysin ennalta määräytynyt, ei sisällä satunnaisuutta
diskonttaus	tulevan arvon muuntaminen nykyarvoksi
dropout	satunnainen neuronien poistaminen mallin koulutuksen aikana mallin suorituskyvyn parantamiseksi
EBIT	liikevoitto ennen korkoja, veroja ja tilinpäätössiirtoja
ekstrapolointi	arvojen arviointi tunnetun havaintoalueen ulkopuolelle
ennustehorisontti	ajanjakso, jolle malli tuottaa ennusteen
epookki	yksi täydellinen koulutuskierron, jonka aikana malli käsittelee koko opetusaineiston
fold	datan yksittäinen osajoukko, joka toimii vuorollaan validointijoukkona ristiinvalidoinnissa
gradientti	funktion suurimman muutosnopeuden suunta ja suuruus
hyperparametri	ennen mallin koulutusta asetettava konfiguraatiomuuttuja, joka ohjaa mallin koulutusta tai rakennetta
IMF	signaalin yksittäinen värähtelykomponentti
inflaatio	rahan ostovoiman yleinen heikkeneminen
interpolointi	tuntemattomien arvojen laskeminen tunnettujen datapisteiden välille
iterointi	toistuva suoritus, edellinen tulos syötetään seuraavaan
jäännösarvo	investoinnin arvo investointiajanjakson päätyttyä

kannattavuus	tuottojen ja kustannusten erotus, voitontuottamiskyky
kantafunktio	joukko yksinkertaisia funktioita, joiden painotettuna summana monimutkaisempi funktio voidaan esittää
KHI	kuluttajahintaindeksi, mittari, joka kuvaa kotitalouksien kulutushyödykkeiden hintojen kehitystä
komposiittimalli	malli, jossa ennuste muodostetaan yhdistämällä useita erikseen estimoituja komponentteja
koneoppiminen	algoritminen menetelmä, jossa malli oppii datasta tekemään ennusteita
korkoa korolle	kertynyt korko liitetään pääomaan, jolloin se alkaa itse tuottaa korkoa
korrelaatio	kahden muuttujan välinen riippuvuus, jota kuvaa korrelaatiokerroin r välillä $-1 \leq r \leq 1$
kovariaatti	selittävä apumuuttuja, joka voi vaikuttaa vasteeseen mutta ei ole itse tutkimuksen pääkohde
kovarianssi	mittaa kahden muuttujan yhteisvaihtelua ja suuntaa
kustannusvastaavuus	periaate, jonka mukaan toiminnan tuottamien tulojen tulee kattaa toiminnasta aiheutuvat kustannukset
kvartaali	vuosineljännes, kolmen kuukauden pituinen ajanjakso
lookback-ikkuna	ennustamisessa käytettävä historiallinen havaintojakso
loss-funktio	häviöfunktio, joka mittaa koneoppimismallin ennustusten ja todellisten arvojen välistä eroa
luokka	olio-ohjelmoinnin käsite, olioiden malli, joka määrittelee niiden ominaisuudet ja toiminnot
MAE	keskimääräinen absoluuttinen virhe
MASE	skaalattu absoluuttinen virhe
Monte Carlo -simulaatio	laskennallinen menetelmä, jossa toistuvilla satunnaisotoksilla arvioidaan tilastollista jakaumaa
MSE	virheiden neliöiden keskiarvo
N-BEATS	syväoppimisen arkkitehtuuri aikasarjaennustamiseen
neuroni	neuroverkon laskentayksikkö, joka muuntaa painotetut syötteet ulostuloksi aktivointifunktion avulla
nimellisarvo	arvo, jota ei ole oikaistu inflaation vaikutuksella, reaaliarvon vastakohta
normalisointi	muuttujien skaalaus yhteismitalliseen vaihteluväliin

NPM	nettovoittomarginaali, mittaa, kuinka suuri osa liikevaihdosta jää tilikauden voitoksi
olio	tiedon ja toiminnan yhdistävä ohjelman osa
omaisuudenhallinta	omaisuuden suunnittelun, hankinnan, kunnossapidon ja uudistamisen systemaattinen ohjaus
omaisuus	liiketaloudessa organisaation taloudelliset resurssit
parametri	mallin sisäinen muuttuja, jonka arvo opitaan datasta
persentiili	prosenttiosuus havainnoista tietyn arvon alapuolella
poisto	hankintamenon kirjaaminen kuluksi usealle tilikaudelle
prolongointi	tulevan arvon laskeminen
PWLF	paloittain määritelty lineaarinen funktio
R ² -luku	mallin selittämä selityskerroin vaihtelusta välillä 0...1
reaaliarvo	inflaatiolla oikaistu arvo, nimellisarvon vastakohta
residuaali	virhetermi, havainnon arvon erotus regressiosuorasta
ristiinvalidointi	menetelmä koneoppimismallin yleistyskyvyn arviointiin
RMSE	neliöllisen keskivirheen neliöjuuri
ROE	oman pääoman tuotto
ROTA	kokonaispääoman tuotto
saneeraus	omaisuuden kuntoa ylläpitävä rakenteiden uusiminen
simulaatio	todellisen ilmiön tai järjestelmän jäljittely mallin avulla
skenaario	hahmotelma tulevan kehityskulun ennakoimiseksi
SMAPE	symmetrinen keskimääräinen prosentuaalinen virhe
spread	kahden eri koron tai osto- ja myyntihinnan välinen ero
stokastinen	satunnaisuutta sisältävä ilmiö tai prosessi
syväoppiminen	koneoppimisen osa-alue, joka perustuu monikerroksiin neuroverkkoihin
tase	raportti yrityksen taloudellisesta tilasta, mittaa yrityksen omaisuuden ja velkojen arvoa tietyssä ajankohtana
tuloslaskelma	mittaa yrityksen tuottoja, kuluja sekä voittoa tai tappiota tietyn ajanjakson aikana
validointi	mallin käyttökelpoisuuden arviointi
vastemuuttuja	muuttuja, jota malli pyrkii ennustamaan
verosuoja	verotuksellinen etu, joka syntyy vähennyskelpoisten erien kuten poistojen ja korkojen kautta
ylisovittaminen	malli mukautuu koulutusdataan liikaa eikä yleisty

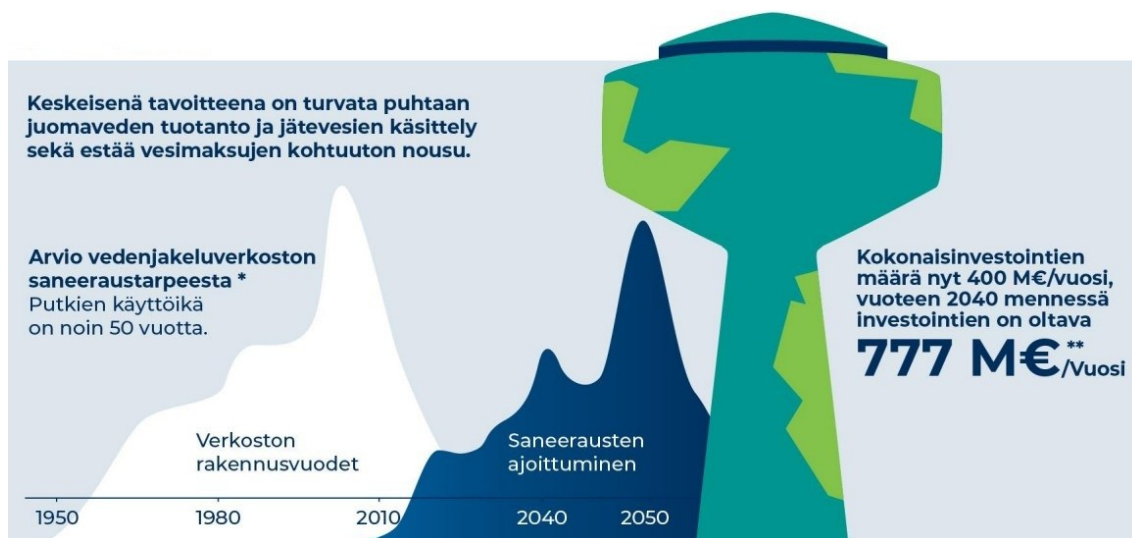
1 JOHDANTO

Suomen vesihuoltosektori on merkittävän rakenteellisen muutoksen keskellä. Maa- ja metsätalousministeriön johdolla valmisteltu kansallinen vesihuoltouudistus johti vesihuoltolain muutokseen (1087/2025), joka tuli voimaan 1.1.2026. Uudistuksella toimeenpannaan pääministeri Petteri Orpon hallitusohjelman kirjausta vesihuollon toiminnan turvaamisesta sekä Vesi on meidän -kansalaisaloitteeseen perustuvaa eduskunnan päätöstä (HE 40/2025 vp, 2025, s. 5). Julkisomisteisten vesihuoltotoimintojen säilyttämisen ja alueellisen yhteistyön edistämisen lisäksi uudistuksen keskeisinä tavoitteina ovat vesihuoltolaitosten omaisuudenhallinnan parantaminen, toiminnan suunnitelmallisuuden lisääminen sekä talouden ja toiminnan valvonnan tehostaminen. Maksujen tulee kattaa vesihuollon kustannukset kestävästi pitkällä aikavälillä, ja niiden määräytymisperusteiden tulee olla läpinäkyviä (HE 40/2025 vp, 2025, s. 22).

Vesihuoltolain uudistus vastaa tilanteeseen, jossa omaisuudenhallintakäytännöt vaihtelevat merkittävästi laitosten välillä eikä systemaattista suunnittelua ole kaikilla laitoksilla lainkaan. Investointeja ja resursseja ei ole suunniteltu riittävän laajasti toiminnan pitkäjänteisyyden varmistamiseksi. Verkoston huono kunto on jätetty huomiotta maksuja asetettaessa, minkä seurauksena kustannusvastaavuus ei toteudu riittävällä tasolla (HE 40/2025 vp, 2025, s. 18, 21). Tilanne korostuu pienissä laitoksissa, joiden henkilöstö- ja talousresurssit ovat riittämättömät suunnitelmalliseen toimintaan. Heikosti resursoitujen laitosten suuri määrä on arvioitu yhdeksi keskeisimmistä rakenteellisista syistä vesihuoltosektorin haasteisiin. Lakiuudistuksella pyritään korjaamaan tämä rakenteellinen puute velvoittamalla kaikki vesihuoltolaitokset systemaattiseen omaisuudenhallinnan suunnitteluun.

Vesihuolto-omaisuus on pitkäikäistä, mutta verkostot ja puhdistamot vanhenevat, ja niiden arvon ylläpito edellyttää kasvavaa investointitasoa (HE 40/2025 vp, 2025, s. 18). Reaktiivinen kunnossapito, jossa toimenpiteisiin ryhdytään vasta vaurioiden ilmettyä, kasvattaa pitkällä aikavälillä sekä korjauskustannuksia että palvelukatkosriskiä. Saneerausvelan taustalla on rakenteellinen kyvykkyyssvaje. Rakennetun omaisuuden tila ROTI 2023 -raportin mukaan vesihuoltolaitosten valmiudet pitkäjänteiseen omaisuudenhallintaan ovat monin osin puutteelliset

(HE 40/2025 vp, 2025, s. 16). Investointien lykkääminen näkyy suoraan kasvavana saneerausvelkana, ja vuotuisen kokonaisinvestointitarpeen on arvioitu nousevan nykytasosta lähes kaksinkertaiseksi vuoteen 2040 mennessä. Kuvio 1 havainnollistaa korjausvelan kehityskulkua.



Kuvio 1. Korjausvelka (Maa- ja metsätalousministeriö, 2025, muokattu). Tiedotteen pääkuva: *Valtioneuvosto (2018), **Suomen Vesilaitosyhdistys ry (2022).

Korjausvelalla tarkoitetaan kumuloitunutta korjausvajetta, jossa vesihuolto-omaisuuden kunnossapito ja saneeraus eivät ole toteutuneet omaisuuden todellisen korjaustarpeen edellyttämässä laajuudessa ja ajoituksessa. Omaisuudenhallintasuunnitelman laadintavelvoite pyrkii katkaisemaan tämän kehityskulun. Kun investointitarpeet ja niiden rahoitus suunnitellaan pitkäjänteisesti, maksukorotukset voidaan toteuttaa maltillisesti vuosien kuluessa sen sijaan, että kertyneet laiminlyönnit purkautuisivat äkillisiksi korotuspaineiksi. Korjausvelan hallinta ja ennakkoiva investointisuunnittelu ovat siten edellytyksiä paitsi vesihuoltolaitosten taloudelliselle kestävyydelle myös yhteiskunnan huoltovarmuudelle.

Suomen vesihuoltojärjestelmä koostuu noin 1 100 vesihuoltolaitoksesta, joiden koko ja käytettävissä olevat resurssit vaihtelevat huomattavasti (HE 40/2025 vp, 2025, s. 16). Vesihuoltolaitoksella tarkoitetaan toimijaa, joka vastaa yhdyskunnan vesihuollosta toimittamalla talousvettä tai vastaanottamalla jätevettä vähintään 10 m³ vuorokaudessa tai vähintään 50 henkilön tarpeisiin. Merkittävä osa laitoksista on pieniä organisaatioita, joilla on rajalliset mahdollisuudet hoitaa toimin-

taansa pitkäjänteisesti ja ammattitaitoisesti sekä varmistaa palvelujen tuottamiseen tarvittavat henkilö- ja taloudelliset resurssit. Pienissä laitoksissa omaisuudenhallintasuunnitelman laatiminen lisää tarvetta ulkopuolisille asiantuntijoille. Suuremmilla laitoksilla on enemmän henkilöstöä ja taloudellisia resursseja, joten ne voivat laatia suunnitelman pääosin itse, mutta myös niiden on kehitettävä dokumentointia ja pitkän aikavälin suunnittelua vaatimusten täyttämiseksi.

Vesihuoltolaki painottaa suunnitelmallista kunnossapitoa, investointien ennakkointia ja talouden läpinäkyvyyttä. Näiden vaatimusten täyttäminen edellyttää laitoksilta aiempaa järjestelmällisempää tiedonhallintaa, mikä luo perustan digitaalisten työkalujen ja tietojärjestelmien hyödyntämiselle omaisuudenhallinnan suunnittelussa. Hallituksen esityksessä todetaankin, että omaisuudenhallintasuunnitelma voidaan laatia tarkoitukseen soveltuvien tietojärjestelmäsovellusten tai -ohjelmistojen avulla (HE 40/2025 vp, 2025, s. 85). Tämä jättää laitoksille liikkumavaraa hyödyntää teknologiaa omaisuudenhallinnan toteuttamisessa omien resurssiensa ja tarpeidensa mukaisesti.

Digitaaliset menetelmät tarjoavat vesihuoltolaitoksille uusia välineitä omaisuudenhallinnan suunnittelun tueksi. Erityisesti koneoppiminen mahdollistaa sanearaustarpeiden tunnistamisen ennen häiriöiden realisoitumista, mikä siirtää omaisuudenhallinnan painopistettä reaktiivisesta korjaamisesta ennakkoivaan riskienhallintaan ja tukee siten lainsäädännön edellyttämää pitkän aikavälin investointisuunnittelua. Pienissä laitoksissa, joissa asiantuntijaresurssit ovat rajalliset, tietojärjestelmäpohjaiset ratkaisut voivat keventää suunnitelman laadintaan liittyvää hallinnollista taakkaa ja parantaa suunnitelmien laatua. Digitaalisten menetelmien hyödyntäminen vesihuoltosektorilla on kuitenkin toistaiseksi ollut epätasaista, ja laitosten välillä on merkittäviä eroja teknologisen valmiuden suhteen. Tämän epätasaisuuden vuoksi tarvitaan ratkaisuja, jotka madaltavat teknologian käyttöönottoa erikokoisille laitoksille, ja tähän tarpeeseen opinnäytetyö pyrkii vastaamaan.

2 TUTKIMUKSEN TIETOPERUSTA

2.1 Tutkimuksen tavoitteet

Tutkimuksen lähtökohtana on vesihuoltolaitosten haaste tuottaa lakisääteisiä suunnitelmia tehokkaasti ja laadukkaasti nykyisillä resursseilla ja menetelmillä. Suunnitteluprosessit ovat usein työläitä, ja ne vaativat sellaista erityisosaamista, jota etenkin pienemmillä laitoksilla ei ole riittävästi käytettävissä. Asiantuntijapalveluiden hankkiminen ulkopuolelta nostaa suunnitelman laadinnan ja ylläpidon kokonaiskustannuksia, mikä rasittaa laitosten taloutta.

Vesihuoltolaitosten omaisuudenhallintaa on tutkittu laajasti. Mutta laajan käsitteen sisältöä ei ole kyetty määrittelemään tyhjentävästi. Myöskään lainsäädäntö ei anna omaisuudenhallintasuunnitelmasta yhtenäistä määritelmää, vaan vastuu sen sisällöstä, muodosta ja toteutuksesta jakautuu eri säädöstasoille. Vesihuoltolain 9 a §:ssä säädetään laitoksen velvollisuudesta laatia omaisuudenhallintasuunnitelma, ja sen vähimmäisisältö sekä vähintään kolmen vuoden päivitysväli on säädetty erikseen valtioneuvoston asetuksen vesihuollosta (1173/2025) 3 §:ssä. Olemassa olevat omaisuudenhallintasuunnitelmat kuvaavat tyypillisesti, mitä hyvä omaisuudenhallinta tulisi pitää sisällään, mutta konkreettisia työkaluja niiden jatkuvaan ylläpitoon ei ole juurikaan esitetty. Tutkimus pyrkii tarjoamaan automatisoidun ratkaisun omaisuudenhallintaan.

Tutkimuksen päätavoitteena on kehittää koneoppimiseen perustuva malli, joka automatisoi ja tehostaa omaisuudenhallintasuunnitelman laatimista ja päivittämistä. Malli hyödyntää vesihuoltolaitoksen olemassa olevaa dataa ja tuottaa sen pohjalta suunnitelmaelementtejä, jotka täyttävät lakisääteiset vaatimukset. Tämän toteuttaminen edellyttää ohjaavan viitekehyksen määrittelyä, jonka puitteissa teknologinen ratkaisu voidaan suunnitella ja toteuttaa.

Päätavoite jakautuu neljään osatavoitteeseen. Ensimmäinen on tunnistaa omaisuudenhallintasuunnittelun keskeiset tietotarpeet ulkoisen ja sisäisen laskenta-toimen näkökulmasta. Toinen on arvioida koneoppimisen soveltuvuutta tähän

kontekstiin. Kolmas on kehittää konseptuaalinen malli, joka mahdollistaa suunnitteluprosessin automatisoinnin. Neljäs on arvioida mallin toteutettavuutta ja sen tuottamia operatiivisia hyötyjä vesihuoltolaitoksen toiminnassa.

Laajemmassa yhteiskunnallisessa kontekstissa tutkimus vastaa kysymykseen, miten digitalisaatio voi tukea vesihuoltosektorin uudistumista ja parantaa toimintavarmuutta. Vesihuolto on kriittinen infrastruktuuripalvelu, jonka häiriötön toiminta on välttämätöntä yhteiskunnan perustoiminnoille. Tutkimuksen tulokset voivat siten tarjota lisäarvoa sekä teknologisesta että strategisesta näkökulmasta.

2.2 Tutkimusmenetelmät

2.2.1 Suunnittelutieteellinen konstruktio tutkimus

Tutkimus pohjautuu suunnittelutieteellisen konstruktio tutkimuksen viitekehykseen, jonka tavoitteena on kehittää ja arvioida innovatiivinen ratkaisu eli artefakti tunnistettuun käytännön ongelmaan. Konstruktio tutkimuksen ydinajatuksena on, että ymmärrys suunnitteluongelmasta ja sen ratkaisusta syntyy artefaktin rakentamisen ja soveltamisen kautta. Artefakti voi olla rakenne, malli, menetelmä tai ilmentymä eli toimivaksi todennettu järjestelmä. (Hevner ym., 2004, s. 75–80.) Tutkimuksessa artefakti on koneoppimista hyödyntävä ohjelmisto, joka tukee vesihuoltolaitoksen omaisuudenhallintasuunnitelman laadintaa ja ylläpitoa.

Suunnittelutieteellisen konstruktio tutkimuksen suunnittelutyö etenee rakentamisen ja arvioinnin iteratiivisena silmukkana, jossa kullakin kierroksella arvioinnin tuottama palaute parantaa sekä artefaktin laatua että ymmärrystä itse ongelmasta (Hevner ym., 2004, s. 78–79). Suunnittelu on luonteeltaan etsintäprosessi, jossa pyritään löytämään käytettävissä olevin keinoin ratkaisu, joka täyttää tavoitteet ja ympäristön asettamat rajoitteet (Hevner ym., 2004, s. 88). Käytännön tietojärjestelmäongelmiin ei yleensä löydy yhtä parasta ratkaisua, joten suunnittelussa tyydytään ratkaisuihin, jotka toimivat riittävän hyvin (Hevner ym., 2004, s. 89).

Hevner ym. (2004, s. 83) esittävät seitsemän ohjenuoraa suunnittelutieteellisen konstruktio tutkimuksen toteuttamiseksi: artefaktin rakentaminen, ongelman merkityksellisyys, suunnittelun arviointi, tutkimuksen tulokset, menetelmällinen huolellisuus ja tarkkuus, suunnittelu etsintäprosessina sekä tutkimustulosten esittäminen. Tutkimuksessa kvalitatiivinen tarkkuus rakentuu aikasarjamenetelmien, talousanalyysin ja koneoppimismallien matemaattiselle teorialle. Konstruktio tutkimuksen periaatteiden mukaisesti tutkimus tuottaa sekä uutta tutkimustietoa että käytännön hyötyä vesihuoltolaitoksille.

2.2.2 Tutkimusmenetelmien normatiivinen konteksti

Vesihuoltolain muutoksen keskeisinä tavoitteina on parantaa vesihuoltolaitosten omaisuudenhallintaa sekä selkeyttää ja tehostaa vesihuollon toiminnan ja talouden valvontaa (HE 40/2025 vp, s. 22). Tavoitteet konkretisoituvat vesihuoltolain 9 a §:ssä, jonka mukaan vesihuoltolaitos huolehtii omaisuudenhallintasuunnitelman laadinnasta ja ajan tasalla pitämisestä. Säännöksen mukaan suunnitelman tarkoituksena on varmistaa, että laitos kykenee huolehtimaan vesihuollosta ja vastaamaan vesihuollon toimintavarmuudesta niin, että vesihuollon kustannusten kattamiseksi perittävät maksut muodostuvat kohtuullisiksi ja tasapuolisiksi.

Omaisuudenhallintasuunnitelmalla on suora yhteys vesihuollon maksuihin, sillä vesihuoltolain 18 §:n 1 momentin mukaan maksujen on pitkällä aikavälillä katettava omaisuudenhallintasuunnitelmassa esitetyt uus- ja korjausinvestoinnit sekä muut kustannukset. Suunnitelman laadinnassa on lisäksi otettava huomioon vesihuoltolain 5 a §:ssä tarkoitettu kunnan vesihuoltosuunnitelma sekä laitoksen koko, toiminnan laajuus ja omaisuuden määrä. Suunnitelma on tarkistettava ja tarvittavilta osin päivitettävä vähintään kolmen vuoden välein.

Suunnitelman sisällöstä säädetään valtioneuvoston asetuksen vesihuollosta (1173/2025) 3 §:ssä. Sisältövaatimukset jäsentyvät neljäksi toisiinsa kytkeytyväksi osa-alueeksi, jotka on havainnollistettu kuviossa 2. Näitä ovat vesihuoltolaitoksen omaisuutta, kuntoa ja kunnossapitoa koskevat tiedot, vähintään 20 vuo-

den aikajänteelle laadittavat investointi- ja rahoitustarpeet, arvio maksujen nykytasosta ja muutostarpeista sekä näihin tietoihin perustuva riskinarvio laitoksen talouden tilasta.



Kuvio 2. Vesihuoltolaitoksen omaisuudenhallintasuunnitelman sisältövaatimukset (valtioneuvoston asetus vesihuollosta 1173/2025, 3 §).

Maksujen arvioinnissa on otettava huomioon investointitarpeet, liittymäärän kehitys ja muut maksutason kannalta oleelliset seikat. Riskinarvio laaditaan kolmen ensimmäisen osa-alueen pohjalta, mikä korostaa niiden keskinäistä riippuvuutta. Omaisuuden kunto määrittää investointitarpeet, jotka heijastuvat maksuihin ja edelleen laitoksen talouden riskiprofiiliin.

2.2.3 Matemaattinen mallinnus ja simulointi

Tutkimuksen empiirinen ydin perustuu matemaattiseen mallinnukseen ja siihen pohjautuvaan simulointiin. Velten ym. (2024, luku 1.2.2) tarkastelevat mallinnusta ja simulointia kokonaisuutena, jossa reaali maailman ongelmaa kuvataan matemaattisen mallin avulla ja mallin toimivuutta arvioidaan simuloinnin, parametrien estimoinnin ja validoinnin keinoin. Mallinnus on heidän mukaansa luonteeltaan

teleologista eli päämääräsuuntautunutta. Mallin käyttötarkoitus ja konteksti määrittävät sen rakenteen ja yksityiskohtaisuuden tason (Velten ym., 2024, luku 1.2.1).

OmaisuuDENhallintamallin rakentaminen noudattaa Veltenin ym. (2024, luku 1.2.2) esittämää mallinnus- ja simulointikehystä, jossa erotetaan toisistaan ongelman määrittely, järjestelmäanalyysi, mallinnus, simulointi ja validointi. Tässä tutkimuksessa ongelmana on tuottaa vesihuoltolaitokselle lakisääteisiä omaisuuDENhallintasuunnitelman elementtejä. Järjestelmäanalyysissa tunnistetaan ilmiön keskeiset muuttujat ja niiden väliset riippuvuudet teorian ja aineiston perusteella. Velten ym. (2024, luku 1.5.1) nimittävät näitä järjestelmän tilamuuttujiksi ja parametreiksi. Mallinnusvaiheessa muuttujien väliset suhteet muotoillaan matemaattisiksi yhtälöiksi ja algoritmeiksi, jotka muodostavat sekä taloudellisen mallinnuksen että koneoppimis pohjaisen ennustamisen perustan ja jotka toteutetaan ohjelmakoodina.

OmaisuuDENhallintamallin valmistuttua ilmiötä tarkastellaan simuloinnin avulla. Velten ym. (2024, luku 1.2.3) määrittelevät simuloinnin mallin soveltamiseksi, jonka tavoitteena on tuottaa strategioita ongelman ratkaisemiseksi tai vastata järjestelmää koskeviin kysymyksiin. Käytännössä mallin parametreja ja lähtöarvoja muutetaan systemaattisesti, jolloin voidaan analysoida eri skenaarioita ja niiden vaikutuksia tuloksiin. Kehyksen viimeinen vaihe on validointi, jossa mallin tuottamia tuloksia verrataan kokeelliseen tai kirjallisuudesta saatavaan aineistoon mallin luotettavuuden ja tarkkuuden arvioimiseksi (Velten ym., 2024, luku 1.2.2).

2.3 Aineisto ja tietolähteet

Tutkimuksen kohteena oleva case-vesihuoltolaitos on osakeyhtiömuotoinen, pieneksi luokiteltava vesihuoltolaitos, joka huolehtii toiminta-alueensa talous- ja jätevesihuollosta. Laitoksen liikevaihto on viime vuosina ollut noin 1,8 miljoonaa euroa ja taseen loppusumma noin 5,7 miljoonaa euroa. Taseen loppusumma on siten yli kolminkertainen liikevaihtoon nähden. Tämä kuvastaa vesihuollon pääomavaltaisuutta ja pitkäikäiseen omaisuuteen sitoutunutta pääomaa, vaikka suhdeluku jää tyypillisesti suurempien laitosten vastaavaa tasoa pienemmäksi.

Suomessa vesihuoltolaitokset toimivat hyvin eri tavoin järjestäytyneinä, kuten osakeyhtiöinä, kunnan liikelaitoksina, kuntayhtymien osina, kunnan teknisen osaston taseyksikköinä sekä asiakasomisteisina vesiosuuskuntina, mikä vaikuttaa tilinpäätöstietojen julkisuuteen ja saatavuuteen (Tuorila & Saastamoinen, 2022, s. 27–32, 72). Tutkimuksen pääasiallisen lähdeaineiston muodostavat osakeyhtiömuotoisen case-vesihuoltolaitoksen tilinpäätös-, kirjanpito- ja investointitiedot. Tutkimuksen käytössä on laitoksen kahdeksan viimeisimmän tilikauden kirjanpitoaineisto, mikä mahdollistaa keskipitkän aikavälin trendien tarkastelun.

Osakeyhtiömuotoisten laitosten tilinpäätökset löytyvät Patentti- ja rekisterihallituksen kaupparekisteristä, kunnallisten liikelaitosten ja kuntayhtymien luvut sisältyvät kunnan tilinpäätökseen, ja vesiosuuskuntien tiedot ovat saatavilla osuuskuntarekisteristä, joskin pienempien osuuskuntien osalta tiedonsaanti voi olla rajoittunutta. Tilinpäätöstiedot ovat siten pääosin julkisia ja vertailukelpoisia, mikä tekee niistä luontevan perustan koneoppimismallin suunnittelulle. Tarkempi laskentatoimen aineisto, kuten kvartaalitaso kustannustiedot, saadaan case-vesihuoltolaitokselta.

Taloudellisina syötemuuttujina käytetään liikevaihtoa, kustannuksia ja taseen loppusummaa, joita täydennetään inflaatiolla ja korkotasolla makrotaloudellisina apumuuttujina. Liikevaihto kuvaa tulorahoituksen tasoa, kustannukset käyttötalousmenoja ja taseen loppusumma omaisuuden laskennallista kirjanpitoarvoa. Tase vahvistetaan vuosittain tilinpäätöksen yhteydessä ja on omaisuudenhallinnan keskeisin seurattava suure. Inflaatiomuuttujana käytetään Tilastokeskuksen kuluttajahintaindeksiä (Tilastokeskus, 2021). Korkotason kehitystä kuvataan 12 kuukauden Euribor-korolla, jonka aikasarja on saatavilla Euroopan keskuspankin tilastotietokannasta (Euroopan keskuspankki, ei pvm.). Korkotaso vaikuttaa pääomavaltaisen toimialan rahoituskustannuksiin. Inflaation ja korkotason huomioiminen mahdollistaa reaali- ja nimellisarvon muutosten erottamisen toisistaan sekä rahoitusympäristön vaikutusten arvioinnin, mikä on välttämätöntä pitkän aikavälin ennustemalleissa. Yhdessä nämä muuttujat kuvaavat laitoksen taloudellista dynamiikkaa ja omaisuuden arvon kehitystä suhteessa ympäröivään makrotalouteen.

2.4 Tutkimuskysymykset

Tutkimusta ohjaavat kolme keskeistä tutkimuskysymystä:

1. Millaisia tietolähteitä ja -aineistoja koneoppimis pohjaisen omaisuudenhallintamallin kehittäminen edellyttää?
2. Millainen arkkitehtuuri ja menetelmäkehitysprosessi tarvitaan toimivan koneoppimismallin rakentamiseksi?
3. Miten koneoppimista voidaan hyödyntää vesihuoltolaitosten omaisuudenhallintasuunnitelman automatisoidussa laadinnassa siten, että lopputulos täyttää vesihuoltolain vaatimukset?

Tutkimuskysymykset etenevät loogisesti datasta menetelmään ja edelleen menetelmän soveltamiseen. Ensimmäinen kysymys tarkastelee mallin koulutuksen ja validoinnin edellyttämää aineistoa sekä sille asetettavia laatu-, kattavuus- ja ajantasaisuusvaatimuksia. Toinen kysymys keskittyy soveltuviin koneoppimis menetelmiin ja omaisuudenhallinnan kontekstissa tarvittavaan tekniseen arkkitehtuuriin ja kehitysprosessiin. Yhdessä nämä kaksi kysymystä muodostavat tutkimuksen teknisen ja menetelmällisen perustan.

Kolmas tutkimuskysymys yhdistää näkökulmat säädösympäristöön ja tarkastelee, miten koneoppimismallin tuottamat tulokset voidaan muuntaa vesihuoltolain sekä valtioneuvoston asetuksen (1173/2025) edellyttämäksi rakenteiseksi omaisuudenhallintasuunnitelmaksi. Keskeisenä tavoitteena on varmistaa lopputuloksen luotettavuus ja todennettavuus. Suunnitelman taustalla olevan datan ja laskennan tulee olla tulkittavaa, ja dokumentaation sellaista, että suunnitelman sisältö ja muodostumisprosessi ovat tarkastettavissa ja perusteltavissa.

2.5 Työn rajaukset

Vesihuoltolaitoksen omaisuudenhallinta on monitahoinen kokonaisuus, joka kytkeytyy laitoksen tekniseen, taloudelliseen ja hallinnolliseen toimintaan. Tutkimus rajautuu tarkastelemaan vesihuoltolain uudistusta omaisuudenhallintasuunnitelman näkökulmasta ja erityisesti sitä, miten suunnitelman laadintaa voidaan tukea koneoppimismenetelmin. Tarkastelu kohdistuu yksinomaan vesihuoltolaitoksiin,

eikä tutkimuksessa käsitellä muita infrastruktuurialoja, kuten energia- tai liikennejärjestelmiä.

Sisällöllisesti tutkimus keskittyy omaisuudenhallintasuunnitelman lakisääteisiin sisältövaatimukseen, painopisteenä investointi- ja taloussuunnittelu. Kuntotutkimukset sekä huolto- ja kunnossapitosuunnitelmat jäävät tarkastelun ulkopuolelle, sillä ne ovat omaisuudenhallintasuunnitelmaa täydentäviä erillisiä asiakirjoja, joiden laadintaan sovelletaan eri menetelmiä ja tietolähteitä. Lisäksi kustannus-hyötyanalyysit, käyttöliittymäsuunnittelu ja muut toiminnallisen toteutuksen yksityiskohdat rajautuvat työn ulkopuolelle, koska tutkimuksen painopiste on mallin menetelmällisessä kehittämisessä. Vesihuoltolaitoksen salassa pidettäviä tietoja ei esitellä tutkimuksessa, ja myös syvälinen kirjanpitoanalyysi jää tarkastelun ulkopuolelle.

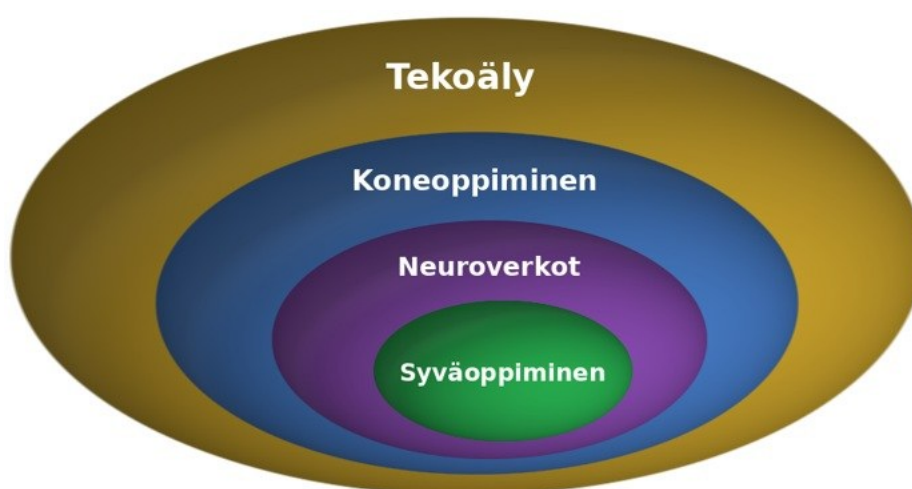
Menetelmällisesti tutkimus sijoittuu syväoppimisen ja ohjatun oppimisen kontekstiin, sillä nämä tarjoavat lupaavia mahdollisuuksia suunnitteluprosessien automatisointiin. Tutkimuksessa keskitytään yhteen valittuun syväoppimismenetelmään. Muiden koneoppimis-, syväoppimis- ja neuroverkkomenetelmien tekninen analyysi rajautuu sen sijaan tarkastelun ulkopuolelle, koska tutkimuksen tavoitteena on rakentaa käsitteellinen kehys yhden valitun menetelmän ympärille eikä vertailla menetelmiä keskenään.

2.6 Keskeiset käsitteet

Vesihuoltolain uudistuksessa omaisuudenhallintasuunnitelma määritellään lakisääteiseksi asiakirjaksi, jonka tarkoituksena on varmistaa vesihuoltolaitosten kyky huolehtia vesihuollosta ja turvata toimintavarmuus pitkällä aikavälillä. Omaisuudenhallinta on systemaattinen prosessi, jonka tavoitteena on ohjata omaisuuden suunnittelua, hankintaa, kunnossapitoa ja uudistamista siten, että saavutetaan haluttu palvelutaso elinkaarikustannukset optimoiden (Paavilainen, 2019, s. 2). Pitkäjänteinen omaisuudenhallinta tukee samalla vesihuoltolaitoksen taloudenhallintaa ja palveluiden toimintavarmuutta. Vesihuolto-omaisuudella tarkoitetaan vesihuoltolaissa määriteltyä infrastruktuuria, joka kattaa vedenkäsittelyyn,

veden jakeluun sekä jätevesien keräämiseen ja käsittelyyn liittyvät rakenteet, laitteet ja järjestelmät. Tässä tutkimuksessa omaisuudella tarkoitetaan vesihuoltolaitoksen kaikkea varallisuutta, mukaan lukien aineettomat resurssit kuten tietojärjestelmät.

Koneoppiminen määritellään tässä tutkimuksessa algoritmiseksi lähestymistavaksi, joka mahdollistaa datasta oppimisen ja ennusteiden tekemisen ilman, että jokainen päätössääntö ohjelmoitaisiin käsin. Kuvio 3 havainnollistaa käsitteiden hierarkkisen suhteen: koneoppiminen on tekoälyn osa-alue, ja syväoppiminen puolestaan koneoppimisen neuroverkkoihin perustuva menetelmäperhe (Sumathi ym., 2022, s. 5).



Kuvio 3. Kone- ja syväoppimisen osa-alueet (Sumathi ym., 2022, s. 6, muokattu).

Koneoppimismenetelmien soveltaminen omaisuudenhallintaan tarjoaa lupaavia mahdollisuuksia erityisesti investointi- ja taloussuunnittelussa. Ennustemallit voivat analysoida historiallisia kustannustietoja ja kulurakenteita, arvioida tulevia kustannuksia, optimoida investointien aikataulutusta ja tukea maksupolitiikan suunnittelua kustannusvastaavuuden varmistamiseksi. Automaatiolla ei tässä yhteydessä tarkoiteta täysin autonomista järjestelmää, vaan teknologista ratkaisua, joka automatisoi rutiinotoimintoja ja tarjoaa analyttistä tukea päätöksenteolle. Inhimillinen toimijuus säilyy keskeisenä erityisesti kriittisissä investointipäätöksissä ja riskienhallinnassa.

3 OMAISUUDENHALLINNAN TALOUSHALLINNUS

3.1 Aikasarja-analyysi

3.1.1 Stokastiset prosessit ja aikasarjat

OmaisuuDENhallinnassa seurataan erityisesti investointikustannusten, omaisuus-erien jälleenhankinta-arvojen sekä käyttö- ja ylläpitokustannusten muutoksia ajan kuluessa. Näiden ilmiöiden analysointi perustuu aikasarjamenetelmiin, joiden matemaattinen perusta nojaa stokastisen prosessin käsitteeseen. Ryan ym. (2025, s. 557) mukaan stokastinen prosessi X määritellään satunnaismuuttujien X_t jakaumien kokonaisuutena, joka on järjestetty aikaindeksin $t \in \{1, 2, \dots\}$ mukaan. Kukin satunnaismuuttuja kuvaa tarkasteltavan suureen mahdollisia arvoja tietyllä ajanhetkellä sekä niiden todennäköisyyksiä.

Stokastisesta prosessista voidaan poimia yksi konkreettinen arvosarja, jolloin kyseessä on prosessin realisaatio eli havaittu aikasarja on

$$X_i = X_1, X_2, \dots, X_n, \quad (1)$$

missä n on aikasarjan pituus. Koska satunnaismuuttujan arvo määräytyy jokaisella ajanhetkellä sen taustalla olevasta jakaumasta, sama prosessi voisi periaatteessa tuottaa lukemattomia erilaisia realisaatioita. Kaikkien näiden mahdollisten realisaatioiden muodostama kokonaisuutta Ryan ym. (2025) kutsuvat prosessin ensembleksi. He esittävät ensemblen kahdella toisiaan vastaavalla tavalla. Ensimmäisen määritelmän mukaan ensemble on kaikkien sellaisten aikasarjojen joukko, jotka prosessi voisi tuottaa. Toisen määritelmän mukaan se on satunnaismuuttujien X_t jakaumien kokonaisuus ajan kuluessa, koska kullakin ajanhetkellä prosessilla on oma jakaumansa, ja näiden jakaumien yhdistelmä määrittää ensemblen. Yhteys käy ilmi tarkastelemalla yksittäistä ajanhetkeä t , sillä jos kerättäisiin kaikkien mahdollisten realisaatioiden arvot tällä ajanhetkellä, muodostuisi täsmälleen satunnaismuuttujan X_t perusjoukon jakauma. (Ryan ym., 2025, s. 557–558.)

Empiirisessä tutkimuksessa on kuitenkin tavallista, että käytettävissä on vain yksi realisaatio, vaikka tilastollisen päättelyn kohteena on koko ensemblen kuvaama prosessi. Tämä asetelma asettaa lähtökohtaisia rajoitteita havaitusta aineistosta tehtäville päätelmille. Stationaarisuusoletukset ovat siksi keskeisessä asemassa, kun yksittäisestä aikasarjasta pyritään tekemään yleistettäviä tilastollisia päätelmiä. (Ryan ym., 2025, s. 557–558.)

Stationaarisuus on stokastisen prosessin ominaisuus, joka kuvaa sen jakaumaominaisuuksien pysyvyyttä ajan suhteen. Aikasarja on ensimmäisen asteen stationaarinen, jos satunnaismuuttujan X_t odotusarvo on vakio kaikilla ajanhetkillä t , eli

$$\mathbb{E}(X_t) = \mu. \quad (2)$$

Jos lisäksi prosessin autokovarianssifunktio

$$\gamma_X(h) = \mathbb{E}[(X_{t+h} - \mu)(X_t - \mu)], \quad (3)$$

missä h on aikaviive, on riippumaton ajanhetkestä t , aikasarjaa kutsutaan toisen asteen stationaariseksi. Mikäli $\gamma_X(h)$ on lisäksi äärellinen, prosessi on heikosti stationaarinen. Vahva stationaarisuus on stationaarisuuden tiukin muoto, mikä edellyttää, että prosessin kaikki tilastolliset ominaisuudet säilyvät samoina ajan suhteen. Stationaarisuus on keskeinen oletus aikasarjojen tilastollisessa päättelyssä, sillä monet estimaattien, luottamusvälien ja testien teoreettiset ominaisuudet perustuvat tähän oletukseen. (Ryan ym., 2025, s. 558.)

3.1.2 Aikasarjan dekomponointi

Aikasarjan dekomponointi on analyyttinen menetelmä, jonka tarkoituksena on syventää aikasarjadatan rakenteellista ymmärrystä hajottamalla havaittu sarja erillisiin komponentteihin: tasoon, trendiin, kausivaihteluun ja satunnaisvaihteluun. Menetelmä tarjoaa tulkinnan mallinnuksen kompleksisuudesta ja soveltuvista lähestymistavoista kunkin komponentin tarkkaan kuvaamiseen, sillä eri komponentit voivat edellyttää toisistaan poikkeavia mallinnusstrategioita (Lewinson, 2022, s. 144). Aikasarjan dekomponointia voidaan soveltaa joko esikäsitteilyvaiheessa, jolloin trendikomponentti poistetaan stationaarisuuden saavuttamiseksi ennen

jäljelle jäävän sarjan mallintamista, tai osana varsinaista mallia, jolloin algoritmillemme tarjotaan riittävästi dataa tai piirteitä trendin oppimiseksi suoraan.

Aikasarjan komponentit jaetaan systemaattisiin ja ei-systemaattisiin. Systemaattiset komponentit ovat säännönmukaisia ja matemaattisesti mallinnettavissa. Näihin kuuluvat taso, trendi ja kausivaihtelu. Taso kuvaa aikasarjan keskimääräistä arvoa, trendi sarjan pitkäaikaista muutossuuntaa peräkkäisten ajanhetkien välillä ja kausivaihtelu keskiarvosta toistuvasti poikkeavaa vaihtelua, joka noudattaa tunnetun pituisia syklejä. Ei-systemaattinen komponentti on kohina, joka edustaa aikasarjan häiriömuuttujien satunnaisvaihtelua ja koostuu vaihteluista, jotka jäävät jäljelle systemaattisten komponenttien poistamisen jälkeen. Kohinaa ei voida mallintaa deterministisesti. (Lewinson, 2022, s. 144.)

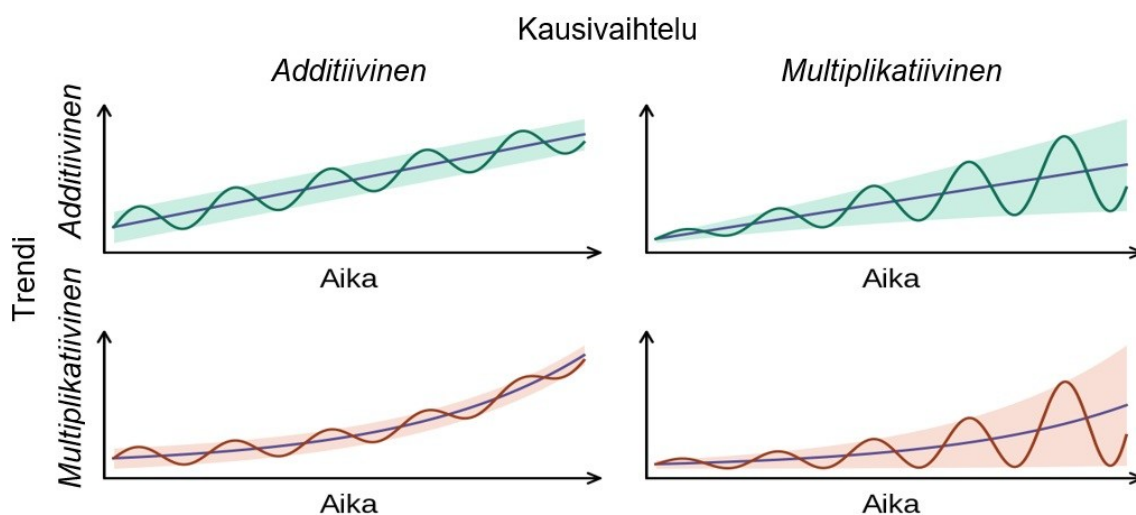
Komponenttirakenteen ymmärtäminen tukee ennustemallien rakentamista ja arviointia. Jos malli ei onnistu erottamaan trendiä ja kausivaihtelua toisistaan, sen residuaaleihin eli havaittujen ja ennustettujen arvojen erotuksiin jää säännönmukaisia rakenteita, jotka paljastavat mallin puutteet. Lewinson (2022, s. 144–145) kuvaa kaksi klassista tapaa yhdistää aikasarjan komponentit: additiivisen ja multiplikatiivisen mallirakenteen. Additiivisessa mallissa havaittu arvo $y(t)$ muodostuu komponenttien summana

$$y(t) = L(t) + T(t) + S(t) + \varepsilon(t), \quad (4)$$

missä $L(t)$ on taso, $T(t)$ trendi, $S(t)$ kausivaihtelu ja $\varepsilon(t)$ kohina ajanhetkellä t . Malli olettaa, että muutokset ovat ajassa vakiosuuruisia siten, että trendi on lineaarinen ja kausivaihtelun amplitudi sekä taajuus pysyvät vakioina. Multiplikatiivisessa mallissa komponentit sen sijaan yhdistyvät kertolaskun kautta

$$y(t) = L(t) \times T(t) \times S(t) \times \varepsilon(t). \quad (5)$$

Rakenne sallii, että kausivaihtelun voimakkuus kasvaa tai pienenee aikasarjan tason mukana, ja mahdollistaa siten epälineaarisen dynamiikan kuvaamisen. Käytännössä aikasarjat voivat sisältää molempia piirteitä samanaikaisesti. Kuviossa 4 esitetään trendin ja kausivaihtelun additiivisten ja multiplikatiivisten yhdistelmien neljä vaihtoehtoa. Käytännössä aikasarjat voivat sisältää näitä piirteitä samanaikaisesti, esimerkiksi additiivinen trendi yhdistettynä multiplikatiiviseen kausivaihteluun.



Kuvio 4. Trendin ja kausivaihtelun additiiviset ja multiplikatiiviset muodot (Lewinson, 2022, s. 145, muokattu).

3.1.3 Inflaatio

Inflaatiolla tarkoitetaan yleisen hintatason nousua ja siitä seuraavaa rahan ostovoiman heikkenemistä. Suomessa hintatason muutosten keskeisin mittari on kuluttajahintaindeksi (KHI). Sen rinnalla käytetään elinkustannusindeksiä ja Euroopan unionin maissa vertailukelpoista yhdenmukaistettua kuluttajahintaindeksiä (YKHI). Indekseillä on osin eri käyttötarkoitukset: etuuksien indeksointi, sopimuskäyttö ja rahapolitiikan inflaatiotavoitteen seuranta. (Tilastokeskus, 2021, s. 12.)

Hintakehitystä kuvataan ensisijaisesti indeksien vuosi- ja kuukausimuutoksina. Tavarat ja palvelut luokitellaan eCOICOP-järjestelmän mukaisesti, mikä tukee vertailukelpoisuutta kansallisesti ja kansainvälisesti. Laskenta perustuu laajaan, kuukausittain kerättävään hintatietoaineistoon ja kotitalouksien todelliseen kuluutukseen pohjautuvaan kulutuskoriin. Käytössä on kiinteäpainotteinen hintaindeksi, jonka kulutusrakennetta päivitetään vuosittain ajantasaisuuden parantamiseksi ja mittausharhan vähentämiseksi. (Tilastokeskus, 2021, s. 12–13.)

Vaikka indeksit täyttävät kansainväliset suositukset, niihin liittyy tunnettuja rajoituksia. Kulutustottumusten muutokset aiheuttavat niin sanottua substituutioharhaa, ja yksittäisen kotitalouden kokema inflaatio poikkeaa usein keskimääräisestä kehityksestä. (Tilastokeskus, 2021, s. 14–15.) Rajoituksista huolimatta KHI muodostaa vakiintuneen käytännön inflaation seurannalle.

Inflaation merkitys ulottuu pelkkää hintojen seurantaan pidemmälle, koska se vaikuttaa olennaisesti rahan aika-arvoon ja siten vesihuoltolaitoksen talous- ja investointilaskelmiin. Knüpfer ja Puttonen (2018) selittävät, että rahan aika-arvo määräytyy osin inflaation mukaan. Ihmiset arvostavat tulevaisuuden kassavirtoja vähemmän, koska niiden ostovoima on inflaation takia tulevaisuudessa pienempi kuin vastaavien rahasummien tänään. Inflaatio vaikuttaa luonnollisesti investointilaskelmiin. Vuosittain vaihtelevalle inflaatiolle kaava voidaan esittää sekä tiivistetyssä että auki kirjoitetussa muodossa

$$V_n = V_0 \times \prod_{i=1}^n (1 + r_i), \quad (6)$$

$$V_n = V_0 \times (1 + r_1) \times (1 + r_2) \times \dots (1 + r_n), \quad (7)$$

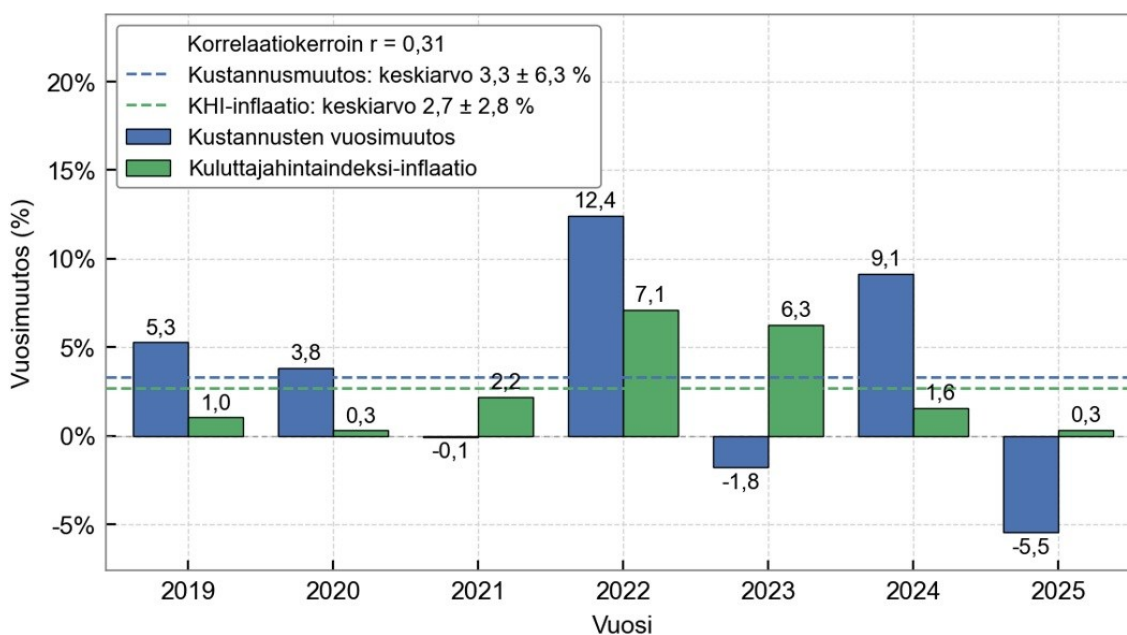
missä V_0 on alkuperäinen arvo, r_i on vuoden i inflaatioprosentti ja V_n on tuleva-arvo n vuoden jälkeen. (Knüpfer & Puttonen, 2018, s. 126.) Π on tulonotaatio eli kaikkien tekijöiden tulo.

Investointilaskelmissa voidaan käyttää joko nimellistä tai reaalista laskentakorkokantaa. Nimellinen korkokanta sisältää inflaatioprosentin. Reaalinen laskentakorko taas on puhdistettu inflaatiosta

$$(1 + i) = (1 + r) \times (1 + \pi), \quad (8)$$

missä i on nimelliskorko, r on reaalikorko ja π on odotettu inflaatio. Kun reaalikorko ja inflaatio ovat arvoltaan pieniä, niiden ristitulo $r \times \pi$ jää käytännössä merkityksettömäksi, jolloin yhtälö yksinkertaistuu muotoon $i \approx r + \pi$. (Nieminen & Nieminen, 2013, s. 335.)

Kuvio 5 havainnollistaa, miten kustannusinflaatio välittyy case-vesihuoltolaitoksen kustannusrakenteeseen. Talousteorian mukaan yleinen hintainflaatio nostaa vesihuoltolaitoksen panoshintoja, mikä kasvattaa kokonaiskustannuksia. Yhteys ei kuitenkaan ole suoraviivainen. Kuvion 5 kohtalainen positiivinen korrelaatio osoittaa, että inflaatio selittää osan kustannusmuutoksista, mutta laitospaikoittaiset tekijät kuten toiminnan laajuus, tehokkuus ja sopimusrakenteet ovat merkittäviä. Kustannukset reagoivat lisäksi viiveellä, mikä selittää vuosien välisiä eroja. Inflaatio tuottaa siten yleistä kustannuspainetta, mutta toteutunut kehitys määräytyy osittain laitoksen omista päätöksistä ja toimintarakenteesta.



Kuvio 5. Case-vesihuoltolaitoksen operatiivisten kustannusten vuosimuutos ja yleinen kuluttajahintojen nousu vuosina 2019–2025.

3.2 Vesihuoltolaitoksen investointi- ja taloussuunnittelu

3.2.1 Vesihuoltolaitoksen omaisuudenhallinta

Omaisuudenhallinnan menettelyjen käyttöönotto edellyttää vesihuoltolaitokselta systemaattista itsearviointia ja kehittämissuunnitelmaa. Paavilainen (2019) esittää kahdeksan peruskysymystä, joihin laitoksen tulisi kyetä vastaamaan omaisuudenhallinnan menettelyjä tarkasteltaessa: mitä omaisuutta laitoksella on, mikä on sen kunto ja arvo, mitä omaisuudella tulisi saavuttaa, mitä sille tulisi tehdä, milloin toimenpiteet tulisi toteuttaa, mikä on niiden kustannus ja miten ne toteutetaan. Näistä viisi ensimmäistä muodostavat perustiedon, jonka tulisi olla teknisen johdon hallussa laitoksen koosta riippumatta. Hyvä omaisuudenhallinta auttaa vertailemaan ja valitsemaan oikeat ratkaisut viimeisiin kolmeen kysymykseen käytettävissä olevien resurssien ja tarjotun palvelun perusteella (Paavilainen, 2019, s. 3.)

Ugarelli ja Sægrov (2022) kuvaavat infrastruktuurin omaisuudenhallinnan historiallista kehitystä kolmena toisiinsa kytkeytyvänä ajanjaksona, joita voidaan tarkas-

tella myös samanaikaisina ja toisiaan vahvistavina kehityssuuntina. Ensimmäinen näkökulma painottaa tiedonkeruuta ja tietojärjestelmien rakentamista, toinen tuottaa riskianalyysiin perustuvia menetelmiä päätöksenteon tueksi, ja kolmas laajentaa tarkastelun kokonaisvaltaiseen kestävyYTEEN, joka ulottuu hallinnollisiin, sosiaalisiin, taloudellisiin ja ympäristöllisiin ulottuvuuksiin. (Ugarelli & Sægrov (2022, s. 3.) Nämä kehityspiirteet ovat tunnistettavissa myös suomalaisessa vesihuoltosektorissa, jossa omaisuudenhallinnan menettelyt rakentuvat samanaikaisesti systemaattisen tiedonkeruun, riskiperusteisen analyysin sekä kestävyyttä korostavan strategisen suunnittelun varaan.

Omaisuudenhallinnan päätöksenteon tulisi perustua omaisuuden elinkaaren arvon muodostumisen arviointiin (Paavilainen, 2019, s. 8). Tämä tarkoittaa käytännössä omaisuuden elinkaarikustannusten ja sillä tuotettavien hyötyjen optimointia. Vesihuolto-omaisuuden arvon muodostuminen rakentuu turvallisuuden, luotettavuuden, ympäristöllisen kestävyuden, tehokkuuden, laadun ja asiakastytyväisyyden vuorovaikutuksesta.

Omaisuuden jälleenhankinta-arvolla tarkoitetaan uuden, vastaavan omaisuuserän todennäköisiä rakennuskustannuksia arviointihetkellä (Paavilainen, 2019, s. 21). Korjausvelka on käsitteenä keskeinen omaisuudenhallinnan mittari, joka kuvaa sitä rahamääräistä arvoa, joka tarvitaan omaisuuserän palauttamiseksi kohtuulliseen kuntoon tai sille asetettuun palvelutasotavoitteisiin liittyvään suorituskykytasoon. Korjausvastuu puolestaan tarkoittaa omaisuuserän palauttamiseksi uutta vastaavaksi tarvittavaa rahamäärää. Nämä käsitteet eroavat toisistaan siten, että optimikuntotaso asetetaan yleensä alle 100 prosentin tasolle, tyypillisesti noin 75 prosenttiin uutta vastaavasta kunnosta. (Paavilainen, 2019, s. 22.)

Infrastruktuurin nykykunnan indeksi IVI kuvaa nykyarvon ja jälleenhankinta-arvon suhdetta. Alegre ja Coelho (2010) esittävät, että infrastruktuurin nykykunnan indeksin optimaalinen arvo on noin 0,5, mikä tarkoittaa, että infrastruktuurin keskimääräinen rapistuminen on puolessa välissä elinkaarta (Alegre & Coelho, 2010, s. 66). Vesihuolto-omaisuuden nykykunnan indeksi voidaan laskea kaavalla

$$IVI = \frac{V_{nyky}}{V_{jh}}, \quad (9)$$

missä V_{nyky} on infrastruktuurin nykyarvo ja V_{jh} on infrastruktuurin jälleenhankinta-arvo. IVI kuvaa rakenteen elinkaaren vaihetta vaihteluvälillä $0 \dots 1$, jossa arvo 1 merkitsee uudenveroista tilaa ja arvo 0 täysin elinkaarensa päässä olevaa rapistunutta rakennetta.

Jälleenhankinta-arvo V_{jh} kuvaa kustannusta, joka syntyisi koko vesihuolto-omaisuuden korvaamisesta uudella nykyhinnoin, ja nykyarvo V_{nyky} huomioi omaisuuserien kulumisen niiden iän ja tavoitepitoajan perusteella. Kun oletetaan yhtenäinen keskimääräinen tavoitepitoaika T , nykyarvo voidaan esittää yksittäisten omaisuuserien summana

$$V_{\text{nyky}} = \sum_{i=1}^n c_i \times q_i \times \left(1 - \frac{t_i}{T}\right), \quad (10)$$

missä c_i on omaisuuserän i yksikkökustannus, q_i sen määrä ja t_i sen ikä. Yhdistämällä kaavat (9) ja (10) indeksi voidaan tulkita kulumiskertoimen kustannuspainotettuna keskiarvona, jolloin se sievenee muotoon

$$\text{IVI} = 1 - \frac{t_m}{T}, \quad (11)$$

missä t_m on omaisuuserien kustannuspainotettu keski-ikä. Indeksillä riippuu siten suoraan omaisuuden ikärakenteesta. Vesihuolto-omaisuuden tavoitteena pidetään välillä $0,5 \dots 0,75$ asettuvaa arvoa, joka kuvaa tasapainoisesti uusiutuvaa ikärakennetta.

3.2.2 Vesihuoltolaitoksen investointisuunnittelu

Investoinneilla tarkoitetaan yleensä pitkäaikaisia ja merkitykseltään suuria taloudellisia hankkeita. Investointeja tulee pystyä arvioimaan osana taloudellista suunnittelua ja päättämään, mihin projekteihin kannattaa ryhtyä ja mihin ei (Knüpfer & Puttonen, 2018, s. 106). Investointipäätösten yhteiskunnallinen konteksti rakentuu kestäväälle kehitykselle. Amadi-Echendu ym. (2010, s. 263) määrittelevät kestävä kehityksen periaatteeksi sen, että nykyisten palvelutarpeiden täyttäminen ei saa vaarantaa tulevien sukupolvien mahdollisuuksia. Investointistrategian onnistuminen edellyttää, että taloudelliset resurssit mitoitetaan tukemaan tavoiteltua saneerausvauhtia.

Verkostot eli vesijohdot ja viemärit muodostavat tyypillisesti valtaosan noin 80 prosenttia vesihuoltolaitoksen käyttöomaisuuden arvosta (Kuulas ym., 2020, s. 63). Tästä seuraa, että investointistrategian painopiste on välttämättä verkosto-omaisuuden hallitussa uusimisessa. Investointiohjelman taloudellinen perusta on omaisuuden jälleenhankinta-arvoon suhteutettu vuotuinen investointitaso. Kun arvioidaan investoinnin kannattavuutta, tehdään ennuste investointihankkeen taloudellisista seuraamuksista (Knüpfer & Puttonen, 2018, s. 107). Tarvittava vuosittainen saneerausaste R johdetaan suoraan verkoston saneerauskierron keskimääräisestä tavoiteajasta T kaavan (12) mukaisesti

$$R = \frac{1}{T}. \quad (12)$$

Kun tavoitteena on uusida koko verkosto-omaisuus 50 vuodessa, saadaan vuosittaiseksi saneerausasteeksi $1/50$, eli karkeasti 2 % kokonaisomaisuudesta vuodessa. Tarvittava vuosittainen saneeraus määrä I_s saadaan kertomalla saneerausaste omaisuuden jälleenhankinta-arvolla V_{jh} kaavan (13) mukaisesti

$$I_s = R \times V_{jh} = \frac{V_{jh}}{T}. \quad (13)$$

Mikäli laitoksen koko teknisen omaisuuden jälleenhankinta-arvo on esimerkiksi 100 miljoonaa euroa, kahden prosentin vuotuinen saneerausaste edellyttää vähintään kahden miljoonan euron vuosittaista investointitasoa. Kun $I_s = R \times V_{jh}$, syntyy tasapainotila, jossa saneerausinvestointien virta kompensoi verkoston rapistumista ja nykykunnan indeksi säilyy tavoitevälillä 0,5...0,75. Jos vuotuinen investointitaso jää tämän alapuolelle, omaisuuskannan ikäjakauma painottuu vanhempiin omaisuuseriin ja nykykunnan indeksi laskee tavoitevälin alarajan 0,5 alapuolelle. Tällöin korjausvelka kasvaa. Mitoitusperiaate skaalataan inflaatio- ja kustannuskehitys huomioiden koko 20 vuoden suunnittelujaksolle.

Investointien suunnittelussa ja investointilaskelmissa merkityksellisiä ovat kassavirrat, eivät suoraan yrityksen tuloslaskelmasta saatavat luvut. Investointien kassavirroista voidaan erottaa kolme päätyyppiä: investointien hankintameno, vuotuiset nettotulot ja investoinnin jäännösarvo. (Niskanen & Niskanen, 2013, s. 324.) Vesihuoltolaitoksen investointien hankintamenoa on usein haastavaa arvioida tarkasti.

Vesihuoltolaitoksen näkökulmasta investointien rahoituskysymykset jakautuvat kahteen osaan taseen mukaisesti. Taseessa vastaavaa-puolella näkyy yrityksen omaisuus (Niskanen & Niskanen, 2013, s. 45). Knüpfer & Puttonen selittävät taseesta, että vastaavaa-puolen kysymykset liittyvät omaisuuden tuottoon eli miten saadaan yrityksen omaisuudelle mahdollisimman hyvä tuotto. Vastattavaa-puolen kysymykset liittyvät hankittavan rahoituksen kustannuksiin, eli miten saadaan rahoitusta mahdollisimman edullisesti ja miten investoinnit rahoitetaan omalla ja vieraalla pääomalla. (Knüpfer & Puttonen, 2018, s. 17.)

Omaisuuksienhallinnan riskiperustaisuutta voidaan havainnollistaa yksinkertaisella vaikutusketjulla, joka kuvaa tiedon ja päätösten etenemistä vaiheittain kuvion 6 mukaisesti. Verkoston kuntotieto muodostaa tietopohjan, jonka varaan saneerauskohteiden priorisointi rakentuu. Priorisointi puolestaan ohjaa operatiivisia toimenpiteitä, ja 20 vuoden investointi- ja taloussuunnitelmat ankkuroivat päätökset taloudelliseen kestävyteen.



Kuvio 6. Investointi- ja taloussuunnittelun sovitus omaisuudenhallinnassa.

3.2.3 Vesihuoltolaitoksen taloussuunnittelu

Vesihuoltolaitoksen kirjanpito kuvaa laitoksen rahaprosessia. Tilikauden aikaisen kirjausten pohjalta laaditaan tilinpäätös, jonka laadintaa säätelevät kirjanpitolaki ja kirjanpitoasetus. Tilinpäätöstietojen analysointiin on kehitetty monenlaisia

tunnuslukuja, jotka voidaan jakaa kolmeen luokkaan analysoitavan kohteen mukaan (Knüpfer & Puttonen, 2018, s. 235): kannattavuuden, tehokkuuden ja vaka-varaisuuden mittareihin. Yleisin kannattavuuden rakennetta kuvaava tunnusluku on liikevoittoprosentti (Paranko, 2012, s. 40), joka määritellään kaavan (14) mukaisesti

$$\text{Liikevoittoprosentti} = \frac{\text{Liiketulos}}{\text{Liiketoiminnan tuotot yhteensä}} \times 100. \quad (14)$$

Tämän opinnäytetyön taloussuunnittelussa keskeinen kysymys on liikevaihdon käsitteen soveltuvuus tutkimusasetelmaan. Paranko (2012) selittää väitöskirjassaan, että kirjanpidollinen liikevaihto edustaa luonteeltaan suppeaan tulokäsitteeseen perustuvaa mittaria, minkä vuoksi se ei välttämättä kata kaikkia tuottoeriä, jotka tutkimuksen näkökulmasta olisi tarkoituksenmukaista sisällyttää tuottojen kokonaisuuteen. Koska tutkimuksessa pääoman mittarina hyödynnetään taseen kokonaissummaa, tuottomuuttujan tulisi periaatteessa sisältää kaikki tuotot, jotta pääoman ja tuottojen välinen tarkastelu olisi sisällöllisesti johdonmukainen. Tätä epäsuhtaa osin lieventää se, että voiton mittarina käytettävä liikevoitto ei myöskään sisällä kaikkia kustannuseriä, jolloin tuottojen ulkopuolelle jäävien erien vaikutus ei välttämättä realisoidu analyysissä täysimääräisesti. (Paranko, 2012, s. 65.)

Vesihuoltolaitosten taloudellinen kestävyys edellyttää pitkäjänteistä suunnittelua, jossa investointitarpeet, kustannusrakenteet ja tuottotavoitteet sovitetaan yhteen. Vesihuoltolain 18 §:n mukaan maksutuloilla on pitkällä aikavälillä katettava uus- ja korjausinvestoinnit sekä kustannukset, ja maksuihin saa sisältyä enintään kohtuullinen tuotto pääomalle. Kohtuullisen tuoton käsite kytkeytyy kokonaispääoman tuottoasteeseen, joka mittaa laitoksen kannattavuutta ja tehokkuutta suhteessa sijoitettuun varallisuuteen. Pääoman kohdentamisen näkökulmasta tämä tarkoittaa, että varat kannattaa ohjata investointeihin niin kauan kuin laitoksella on palvelukyvyyn turvaavaa saneeraustarvetta. Vasta kun investointitarve on katettu, jakokelpoisten varojen palauttaminen omistajalle tulee ajankohtaiseksi.

Yleisesti yrityksen pitkäjänteisen taloussuunnittelun päämääränä on varmistaa omistajien sijoittamalle pääomalle saatava tuotto kestäväällä ja ennustettavalla tavalla. Kokonaispääoman tuotto (engl. return on assets, ROA) ja oman pääoman

tuotto (engl. return on equity, ROE) muodostavat taloudellisen suorituskyvyn arvioinnin perusmittariston, sillä ne ilmentävät yrityksen kykyä tuottaa voittoa suhteessa käytettävissä olevaan varallisuuteen tai omistajien sijoittamaan pääomaan (Fretes, 2025, s. 42). Vesihuoltolaitos ei kuitenkaan pyri maksimoimaan tuottoa omistajalleen. Pelkkä tuottolukujen tarkastelu ei myöskään riitä kestävän arvonluonnin varmistamiseen, vaan suunnittelun tueksi tarvitaan ymmärrys siitä, mistä osatekijöistä tuotto muodostuu ja mihin johdon toimenpiteillä voidaan vaikuttaa. Tähän tarpeeseen vastaa DuPont-analyysi, joka kehitettiin DuPont de Nemours -yhtiössä vuonna 1919 ja on sittemmin vakiintunut keskeiseksi tavaksi dekomponoida tuottomittareita (Galabov & Zlateva, 2019, s. 53–54).

DuPont-analyysin vahvuus perustuu siihen, että se jäsentää ROA:n ja ROE:n matemaattisesti toisiinsa kytkeytyviksi osatekijöiksi, jolloin kunkin osa-alueen vaikutus kokonaistuottoon tulee näkyviin. Erittelyrakenne mahdollistaa täsmällisen analyysin siitä, missä määrin yrityksen kannattavuus selittyy operatiivisella tehokkuudella, pääoman allokoinnilla tai rahoitusrakenteella (Fretes, 2025, s. 42–43; Galabov & Zlateva, 2019, s. 54–55).

3.3 DuPont-analyysi

DuPont-järjestelmän teoria rakentuu taloudellisten tuottosuhteiden algebralliselle dekomponoinnille, jossa kokonaistuottomittarit hajotetaan osatekijöiksi. Oman pääoman tuotto ROE voidaan esittää DuPont-yhtälön mukaisesti kolmikomponenttisena tulona seuraavasti (Fretes, 2025, s. 43)

$$\text{ROE} = \frac{\text{Nettotulos}}{\text{Liikevaihto}} \times \frac{\text{Liikevaihto}}{\text{Taseen loppusumma}} \times \frac{\text{Taseen loppusumma}}{\text{Oma pääoma}}, \quad (15)$$

$$\text{ROE} = \text{NPM} \times \text{TATO} \times \text{EM}. \quad (16)$$

Yhtälön ensimmäinen termi edustaa nettovoittomarginaalia (engl. net profit margin, NPM), toinen termi pääoman kiertonopeutta (engl. total asset turnover ratio, TATO) ja kolmas termi pääomarakennetta kuvaavaa velkavipukerrointia eli rahoitusrakennetta (engl. equity multiplier, EM). Vastaavasti kokonaispääoman tuotto (engl. return on assets, ROA) voidaan ilmaista kaksikomponenttisena tulona (Irons, 2020, s. 169, 171)

$$\text{ROA} = \text{Nettovoittomarginaali} \times \text{Pääoman kiertonopeus}. \quad (17)$$

Dekomponoinnista nähdään, että laitoksen oman pääoman tuotto riippuu liiketoiminnan operatiivisesta kannattavuudesta, varojen käytön tehokkuudesta sekä rahoitusrakenteesta. DuPont-järjestelmän diagnostinen arvo perustuu siihen, että muutos missä tahansa näistä komponenteista, nettovoittomarginaalissa, pääoman kiertonopeudessa tai velkaantumisasteessa, heijastuu suoraan ROE-tunnuslukuun. Fretes (2025) korostaa, että komponenttitason analyysi mahdollistaa tuoton heikkenemisen taustalla olevien tekijöiden tunnistamisen, onko lasku seurausta marginaalien supistumisesta, varojen käytön tehostommuudesta vai liiallisesta velkavivusta (Fretes, 2025, s. 42–43). Tämä puolestaan auttaa yrityksen johtoa kohdentamaan korjaavat toimenpiteet oikeaan osa-alueeseen.

DuPont-hajotelman kukin komponentti kuvaa yrityksen eri toiminnallista osa-aluetta. Zhang (2025, s. 291–294) kertoo, että NPM mittaa operatiivista kannattavuutta. Se ilmaisee, kuinka suuri osuus liikevaihdosta jää nettotulokseksi kulujen jälkeen ja heijastaa siten kustannusrakenteen tehokkuutta sekä hinnoitteluvaihtoehtoja. TATO mittaa varojen käytön intensiteettiä eli sitä, kuinka paljon liikevaihtoa syntyy kutakin taseeseen sitoutunutta euroa kohden (Knüpfer & Puttonen, 2018, s. 238). Irons (2020, s. 171) toteaa, että EM kuvaa yrityksen rahoitusrakennetta eli vieraan pääoman suhdetta omaan pääomaan.

Paranko (2012) korostaa, että vieraan pääoman vipuvaikutukseksi kutsutaan ilmiötä, jossa sellaisen kannattavuuden aikana, jolloin tuotto ylittää vieraasta pääomasta maksetun koron omalle pääomalle, saadaan lisätuottoa vieraasta pääomasta (Paranko, 2012, s. 41). Korkea velkavipu lisää taloudellista riskiä ja kaventaa rahoituksellista liikkumavaraa. Liiallinen velkaantuminen voi heikentää maksuvalmiutta ja asteittain nostaa rahoitusahdingon todennäköisyyttä erityisesti pitkällä aikavälillä. Vieraan pääoman määrän voimakaskaan lisääntyminen ei välttämättä aiheuta taseongelmaa, mikäli taseen toisella puolella on realisoitavissa olevia likvidejä varoja (Knüpfer & Puttonen, 2018, s. 239).

3.3.1 ROTA-hinnoittelumalli taloussuunnittelussa

Vesihuoltolaitoksen maksujen hinnoittelun tulee pitkällä aikavälillä turvata taloudellinen kestävyys ja mahdollistaa riittävät investoinnit. Koska toiminta on julkisen

valvonnan alaista ja voiton tavoittelu rajattua, kannattavuutta arvioidaan kokonaispääoman tuoton näkökulmasta oman pääoman tuoton sijaan. Tässä mallissa käytetään kokonaispääoman tuottoastetta (engl. return on total assets, ROTA), joka määritellään liikevoiton (engl. earnings before interest and taxes, EBIT) suhteena taseen loppusummaan

$$\text{ROTA} = \frac{\text{EBIT}}{A}, \quad (18)$$

missä EBIT on liikevoitto tai -tappio eli liiketulos ennen korkoja ja veroja, ja A on taseen loppusumma eli kokonaisvarallisuus. Liikevoiton ollessa negatiivinen ROTA on negatiivinen. Jos tuotto muuttuu negatiiviseksi, se voidaan rinnastaa kuluihin (Paranko, 2012, s. 45).

ROTA-tunnusluvun käyttö nettotulospohjaisen ROA-laskennan sijaan on perusteltua vesihuoltolaitoksen kontekstissa, koska EBIT eliminoi rahoitusrakenteen sekä verojen vaikutukset ja kuvaa siten puhtaammin laitoksen operatiivista suorituskykyä. Kannattavuusperusteisessa ROTA-hinnoittelussa liikevaihtotavoite johdetaan tavoitellusta kokonaispääoman tuotto prosentista $\text{ROTA}^{\text{tavoite}}$. Ensin määritellään tavoiteltu liikevoitto $\text{EBIT}^{\text{tavoite}}$ prosenttiosuutena taseen arvosta ja sen jälkeen lasketaan tarvittava vuosittainen liikevaihto tämän tuloksen saavuttamiseksi.

Inflaatio on vesihuoltolaitoksen taloudellisessa suunnittelussa erityisen merkittävä tekijä, koska vesihuolto-omaisuuden pitkät taloudelliset pitoajat altistavat laitoksen kustannustason kumulatiivisille muutoksille vuosikymmenten aikajännteellä. Inflaatio aiheuttaa investointien nimellisten kassavirtojen kasvua (Nieminen & Nieminen, 2013, s. 333). Westerlund (1984) pohtii, että käyttöomaisuus edustaa kirjanpidossa historiallista jäännösarvoa, voiton edustaessa aina kyseisen jakson rahanarvoa. Syntyy kysymys siitä, onko oikein verrata tarkastelujakson rahanarvossa esiintyvää voittokäsitettä sellaiseen pääomakäsitteeseen, jonka suurin komponentti perustuu historialliseen rahanarvoon. Inflaation vallitessa on luonnollista korjata kuluvan käyttöomaisuuden arvo vastaamaan ajankohtaista hintatasoa. (Westerlund, 1984, viitattu lähteessä Paranko, 2012, s. 46.)

$\text{ROTA}^{\text{tavoite}}$ -een asettaminen edellyttää reaali- ja nimellistuoton selkeää erottamista. Kirjanpidolliset poistot perustuvat historiallisiin hankintahintoihin, eivätkä

siten huomioi inflaatiota, mikä johtaa poistojen alimitoitukseen suhteessa omaisuuden todellisiin investointitarpeisiin. Nimellis- ja reaalikoron yhteyden esittävä kaava (8) tarjoaa välineen arvioida velkarahoituksen todellista kannattavuutta, kun korkoa korolle -vaikutus huomioidaan reaalisesti.

3.3.2 Liikevaihtotavoitteen matemaattinen malli

Johdetaan malli, jossa tavoiteltu ROTA muunnetaan liikevaihtotavoitteeksi. Mallin lähtökohtana on DuPont-hajotelma kahdella komponentilla

$$\text{ROTA} = \text{PM} \times \text{TATO} = \frac{\text{EBIT}}{R} \times \frac{R}{A}, \quad (19)$$

missä PM on liikevoittomarginaali (liikevoitto/liikevaihto), TATO on pääoman kiertonopeus (liikevaihto/kokonaisvarallisuus), R on liikevaihto ja A on taseen loppusumma. Kun $\text{ROTA}^{\text{tavoite}}$ on asetettu, tavoiteltu liikevoitto saadaan suoraan

$$\text{EBIT}^{\text{tavoite}} = \text{ROTA}^{\text{tavoite}} \times A. \quad (20)$$

Kustannusrakenne oletetaan tarkasteluhetkellä kiinteäksi lyhyellä aikavälillä, jolloin kustannukset C ovat liikevaihdon ja liikevoiton erotus

$$C = R - \text{EBIT}. \quad (21)$$

Liikevaihtotavoite saadaan lisäämällä tavoiteltu liikevoitto kustannuksiin

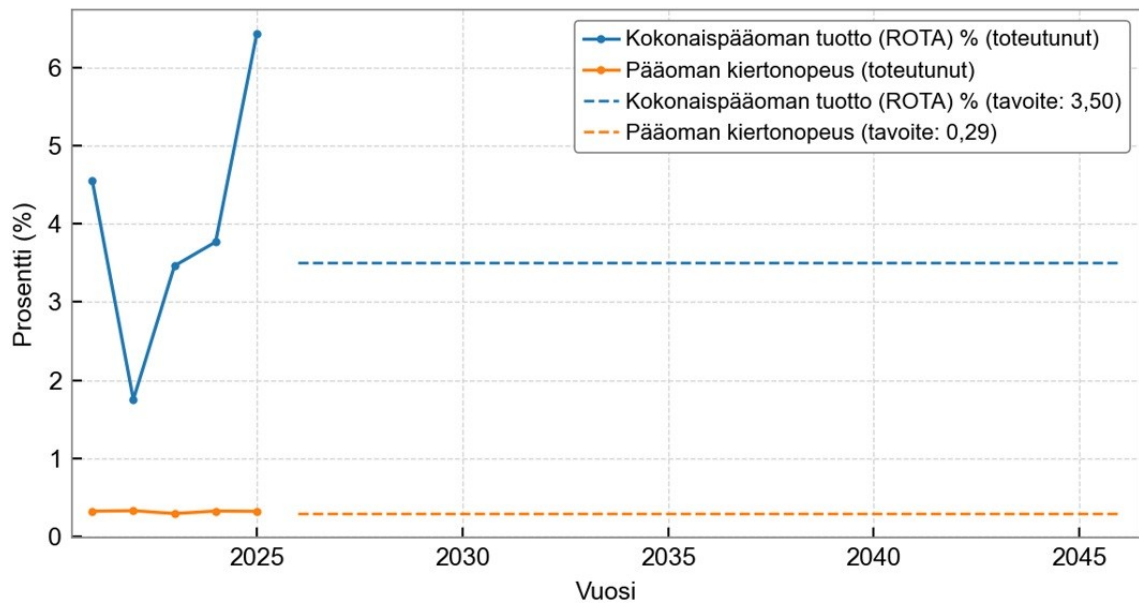
$$R^{\text{tavoite}} = C + \text{EBIT}^{\text{tavoite}}. \quad (22)$$

Tämä voidaan ilmaista myös pääoman kiertonopeuden kautta

$$\text{TATO}^{\text{tavoite}} = \frac{R^{\text{tavoite}}}{A}. \quad (23)$$

Yhtälöt (20)–(23) muodostavat staattisen matemaattisen mallin vesihuoltolaitoksen yhdelle tilikaudelle. Lähtötason $\text{ROTA}^{\text{tavoite}}$ asetetaan laitoksen investointitarpeiden ja toimintaan sidotun pääoman tuotto-odotusten mukaisesti. Knüpferin ja Puttosen (2018, s. 32) mukaan investoinneille asetettavan tuottovaatimuksen tulee perustua investoinnin rahoitukseen käytettävän pääoman kustannuksiin. Huomioitavaa on, että investointiin sitoutuneella pääomalla on aina olemassa kustannus eli niin sanottu vaihtoehtokustannus riippumatta siitä, mistä lähteistä

pääoma on peräisin. Kuviosta 7 nähdään, kuinka case-vesihuoltolaitoksen tuottoasteet vaihtelevat vuosittain, mutta tavoitteet pysyvät ennustehorisontissa vakioina.



Kuvio 7. Case-vesihuoltolaitoksen kokonaispääoman tuottoprosentti ROTA ja pääoman kiertonopeus TATO: historiallinen kehitys ja tavoitetaso.

3.3.3 Kokonaisvarallisuuden keskiarvoanalyysi

Pitkän aikavälin suunnittelussa on huomioitava, että kokonaisvarojen nimellisarvon on kasvettava inflaation vaikutuksesta. Hallitun kasvun taustalla on oletus siitä, että suurin osa taseen eristä kasvaa jatkossakin samassa suhteessa liikevaihtoon (Niskanen & Niskanen, 2013, s. 363). Esitetään hallitusta kasvusta prolongoimalla malli, joka projisoi omaisuuden arvon vuosittaisilla inflaatiotekijöillä. Olkoon A_0 viimeksi vahvistettu tasearvo ja π_t vuoden t inflaatioaste. Kokonaisvarojen laskennallinen nimellisarvo vuonna t saadaan koronkorkperiaatteella

$$A_0 = \text{taseen loppusumma ajanhetkellä } 0, \quad (24)$$

$$A_t = A_{t-1} \times (1 + \pi_t). \quad (25)$$

Liikevaihto kehittyy ajan suhteen, jolloin kokonaisvarat päivitetään kunakin vuonna edellisen vuoden arvosta kertomalla kokonaisvarat inflaation korkotekijällä $(1 + \pi_t)$. Inflaatiomalli ei käytä yhtä vakioinflaatioarvoa, vaan sallii inflaation

vaihdella vuosittain. Koska kokonaisvarat kasvavat pitkällä aikavälillä, myös tarvittava liikevaihto ja -voitto kasvavat. Keskimääräinen kokonaisvarallisuuden A_{ka} arvo N vuoden analyysijaksolla lasketaan kaavalla

$$A_{ka} = \frac{1}{N+1} \sum_{t=0}^N A_t. \quad (26)$$

Tämän perusteella lasketaan keskimääräinen tarvittava liikevoitto $EBIT^{tavoite}$ ja liikevaihto $R^{tavoite}$ kaavoilla (27) ja (28)

$$EBIT^{tavoite} = ROT A^{tavoite} \times A_{ka}, \quad (27)$$

$$R^{tavoite} = C + EBIT^{tavoite}. \quad (28)$$

Yhtälöt (27)–(28) laskevat varallisuuden vuosittaisen kehityksen, keskiarvon ja siitä johdetut tavoitteet ennustetun inflaation perusteella. Malli palauttaa myös suhteellisen tuottovaatimuksen nykyiselle taseelle

$$ROTA_0^{tavoite} = \frac{EBIT^{tavoite}}{A_0}, \quad (29)$$

jolloin tunnusluku kertoo, kuinka suuri tuottoaste nykyiselle taseelle tarvittaisiin, jotta keskimääräisen omaisuuden tuottotavoite täyttyisi koko analyysijaksolla vesihuoltolaitoksen kannattavuuden pysyessä vakiona. Käytännössä kokonaispääoman tuottoprosentin säilyminen vakiona pitkiä aikoja ei ole realistinen oletus, vaan se muuttuu lähes vuosittain. Mutta sen tulee olla pitkällä aikavälillä vähintään tavoitteen verran kokonaisvarallisuuden arvonsäilyttämisen kannalta.

3.3.4 Kustannusten ja liikevaihdon inflaatiodynamiikka

Inflaatiodynamiikan pitkän aikavälin mallissa kustannukset, liikevaihto ja -voitto kasvavat inflaation mukana samalla periaatteella kuin nimellinen kokonaisvarallisuus. Ennustejakson ensimmäisenä vuonna lähtöarvot ovat staattisen mallin yhtälöiden (20)–(22) tuottamat tavoitearvot. Sen jälkeen kukin komponentti päivittyy vuosittaisella inflaatioluvulla

$$C_t = C_{t-1} \times (1 + \pi_t), \quad (30)$$

$$EBIT_t^{tavoite} = EBIT_{t-1}^{tavoite} \times (1 + \pi_t), \quad (31)$$

$$R_t^{tavoite} = R_{t-1}^{tavoite} \times (1 + \pi_t). \quad (32)$$

Nimellinen inflaatiomalli olettaa, että kustannusinflaatio ja tuottojen nousu etenevät samassa suhteessa, jolloin liikevoittomarginaali PM säilyy vakiona ja pääoman kiertonopeus TATO pysyy ennallaan. Tällöin kokonaisvarojen reaaliarvo pysyy muuttumattomana tarkasteluhetkeen nähden. Malli perustuu oletukseen, että lopulta kaikki kustannuserät ovat kokonaisuudessaan muuttuvia kustannuksia pitkällä aikavälillä. Oletus on yksinkertaistava mutta perusteltu pitkän aikavälin suunnittelussa, jossa yksittäisten vuosien poikkeamat tasoittuvat.

Yrityksen taseen vastaavaa-puolen erien oletetaan yleensä käyttäytyvän samalla logiikalla kuin aiemmin tarkastellut kustannuserät. Toisin sanoen myös taseen vastaavaa-puolen eriä ennustetaan suhteessa liikevaihtoon. (Nieminen & Nieminen, 2013, s. 354.) Tämän oletuksen taustalla on ajatus siitä, että yrityksen nykyiset investoinnit eivät riitä tuottamaan ennustettua liikevaihdon kasvua, vaan kasvun tukeminen edellyttää lisäpanostuksia käyttöomaisuuteen ja muihin vastaaviin eriin. Mallissa pääoman kiertonopeus pysyy nimellisarvojen kasvaessa vakiona, koska sekä osoittaja (liikevaihto) että nimittäjä (kokonaisvarallisuus) kasvavat samassa suhteessa

$$\text{TATO}_t = \frac{R_t^{\text{tavoite}}}{A_t} = \frac{R_0^{\text{tavoite}} \times \prod_{i=1}^t (1 + \pi_i)}{A_0 \times \prod_{i=1}^t (1 + \pi_i)} = \frac{R_0^{\text{tavoite}}}{A_0} = \text{TATO}_0. \quad (33)$$

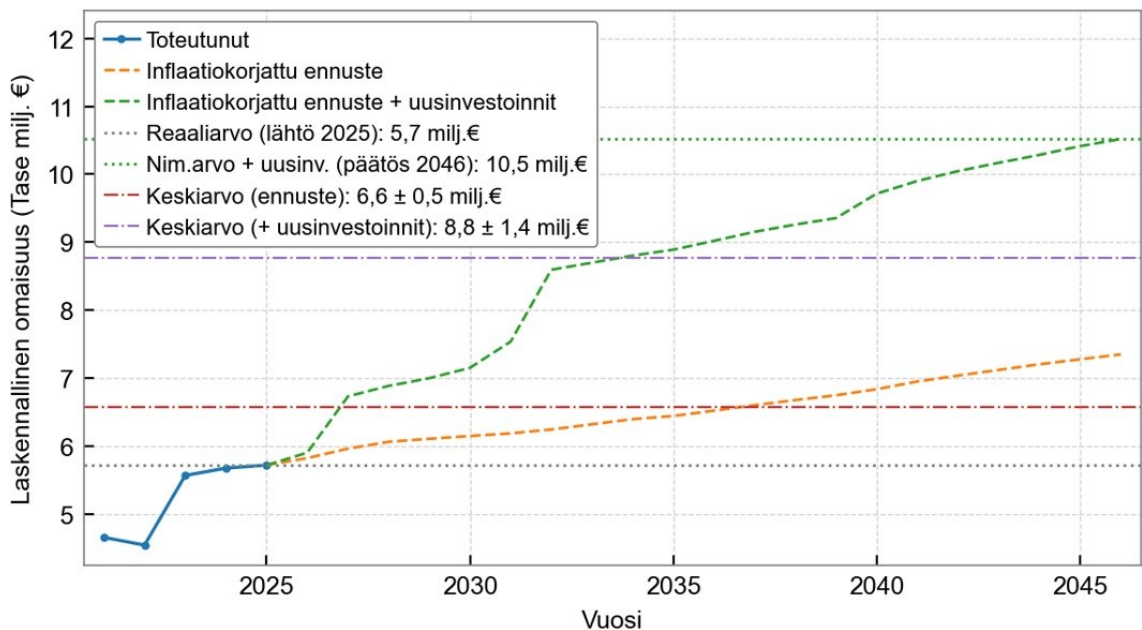
Yhtälöt (30)–(32) kuvaavat inflaatiokorotettua ennustelaskentaa, jossa perusskenaario mallinnetaan ilman uusinvestointeja ja laajennettu skenaario sisältää uusinvestoinnit

$$A_t^u = A_{t-1}^u \times (1 + \pi_t) + I_t^{\text{uusi}}, \quad (34)$$

missä I_t^{uusi} on vuoden t uusinvestointivirta. Molemmissa skenaarioissa kunkin erän arvo korotetaan vuosittaisella inflaatiolla.

Mallin perusoletuksena on, että omaisuuden reaaliarvo pysyy vakiona koko ennustejakson ajan ennen uusinvestointeja (kuvio 8). Kun taseen nimellisarvo kasvaa joka vuosi täsmälleen inflaation verran, se tarkoittaa korkoa korolle -periaatteella sitä, että jokaisen vuoden reaaliarvo on sama kuin lähtövuoden tasearvo. Inflaatio hinnoittelee uudelleen saman fyysisen omaisuuden tarvittavaan uuteen hintatasoon. Todellisuudessa omaisuuden reaaliarvo yleensä laskee ajan myötä kulumisen ja vanhenemisen vuoksi, ellei sitä ylläpidetä saneerauksilla.

Kuvion 8 kokonaisvarallisuuden nimellisarvon kehitys mallinnetaan, jotta voidaan määrittää haluttu liikevoittotavoitteen taso ennustehorisontin funktiona. Parankon (2012, s. 80) mukaan taseen luonnollinen kasvunopeus riippuu liikevaihdon kasvunopeudesta ja liikevoittoprosentista, ”jos liikevoittoprosentti on alhaisempi kuin liikevaihdon kasvuprosentti, niin taseen muutosnopeus alkaa kasvaa kohti liikevaihdon kasvunopeutta. Jos liikevoittoprosentti ylittää liikevaihdon kasvunopeuden, taseen muutosnopeus laskee kohti liikevaihdon muutosnopeutta. Liikevoittoprosentti määrittää sen, onko taseen kasvunopeus hidastuva vai nopeutuva suhteessa liikevaihdon kasvuun”.



Kuvio 8. Case-vesihuoltolaitoksen kokonaisvarallisuuden nimellisarvon ennakoitu kehitys ennustehorisontin funktiona: ennuste inflaatiokorotuksella sekä ennuste uusinvestoinnit huomioiden.

3.3.5 Liikevoittotavoite omaisuuden arvon säilyttämiseksi

ROTA-tavoitteeseen perustuva liikevoittovaatimus ei aina riitä varmistamaan laitoksen investointikykyä. Jos vuosittaiset investoinnit ylittävät poistot, täytyy operatiivisen kassavirran kattaa tämä erotus. Rahoituksen näkökulmasta investointien käsittely kokonaisuudessaan investointien tekohetkellä on perusteltua, koska juuri silloin investointien vaikutukset näkyvät kassassa (Knüpfer ja Puttonen

2018, s. 243). Tämän vuoksi omaisuuden arvonsäilyttämisen malli asettaa liikevoitolle kassavirtaperusteisen vähimmäistason ja yhdistää sen tuottotavoitteen.

Nettoinvestointitarve I_t^{netto} vuonna t on investointien ja poistojen erotus

$$I_t^{\text{netto}} = I_t - D_t \quad (35)$$

missä I_t on saneeraus- ja uusinvestoinnit yhteensä ja D_t on poistot vuonna t . Nettoinvestointitarve ilmaisee, kuinka paljon enemmän tai vähemmän investointeihin tarvitaan rahaa, kuin mitä poistot vapauttavat kassavirtaan. Poistot ovat kirjanpidoollinen kulu, joka ei ole kassasta lähtevää rahaa, vaan ne palauttavat operatiiviseen kassavirtaan aiemmin investoidun summan asteittain omaisuuserän taloudellisen vaikutusajan kuluessa (Knüpfer & Puttonen, 2018, s. 243). Kun nettoinvestointitarve on positiivinen $I_t^{\text{netto}} > 0$, poistot eivät riitä kattamaan investointinevoja ja erotus on rahoitettava liikevoitosta, mikä nostaa kassavirtaperusteista liikevoittoa vaatimusta. Kun investoinnit ja poistot ovat tasapainossa $I_t^{\text{netto}} = 0$, operatiivinen kassavirta kattaa investoinnit ilman lisätuottoa vaatimusta. Kun poistot ylittävät investoinnit $I_t^{\text{netto}} < 0$, kassavirtaa vapautuu yli investointitarpeen eikä kassavirta aseta lisärajoitetta liikevoitolle. Tällöin kassavirtaa voi säästää tuleville investoinneille. Kassavirran (engl. cash flow, CF) turvaava minimi-EBIT on

$$\text{EBIT}(\text{CF}, \text{min})_t = \max(0, I_t^{\text{netto}}). \quad (36)$$

Kun nettoinvestoinnit ovat positiiviset $I_t^{\text{netto}} > D_t$, liikevoiton on oltava vähintään nettoinvestointien suuruinen, jotta operatiivinen kassavirta kattaa investoinnit. Yhtälö varmistaa, ettei negatiivista nettoinvestointia tulkita negatiiviseksi EBIT-vaatimukseksi. Yhdistetty EBIT-tavoite vuonna t on näiden kahden tavoitteen maksimi

$$\text{EBIT}_t^{\text{tavoite}} = \max(\text{ROTA}^{\text{tavoite}} \times A_t, \text{EBIT}(\text{CF}, \text{min})_t). \quad (37)$$

Operatiivinen kassavirta (engl. operating cash flow, OCF) ja vapaa kassavirta (engl. free cash flow, FCF) saadaan

$$\text{OCF}_t = \text{EBIT}_t^{\text{tavoite}} + D_t, \quad (38)$$

$$\text{FCF}_t = \text{OCF}_t + I_t. \quad (39)$$

Tarvittava liikevaihto saadaan lisäämällä yhdistetty EBIT-tavoite kustannuksiin

$$R_t^{\text{tavoite}} = C_t + \text{EBIT}_t^{\text{tavoite}}. \quad (40)$$

Kaava (40) ilmaisee rahoitusrakenteen turvaavan liikevaihdon vuotuisen kasvun omaisuuden reaaliarvon pysyessä vakiona. Kaavat (37)–(39) eivät vastaa kirjanpitolakia. Knüpfer ja Puttonen (2018) kertovat, että liikevoitto kuvastaa yrityksen liikevaihdosta kulujen vähentämisen jälkeen jäljelle jäävää kassavirtaa. Verot vaikuttavat kassaperusteisesti ja siksi ne vähennetään operatiivisesta kassavirrasta. Laskennallisilla veroilla tarkoitetaan liikevoittoa kerrottuna yritysverokannalla. Todellisuudessa maksettujen verojen sijaan käytetään laskennallisia veroja, koska maksetuissa veroissa tuntuu rahoituskulujen veroja vähentävä vaikutus. Tämä vaikutus otetaan huomioon pääoman kustannuksessa, joten tässä sitä ei pidä huomioida. (Knüpfer ja Puttonen 2018, s. 234.) Lisäksi he jatkavat, että liiketoiminnan kassavirta ottaa huomioon kaikki liiketoimintaan liittyvät kassaan ja kassasta maksut. Investointien osalta on siis vielä tehtävä oikaisuja. Yritys voi investoida pääomia kahteen kohteeseen: pitkäaikaisiin investointeihin ja käyttöpääomaan. Nettoinvestoinnit edustavat rahasummaa, jonka yritys on käyttänyt tosiasiallisesti investointeihin. Käyttöpääomaan vaikuttavat positiivisesti vaihto-omaisuus ja myyntisaamiset ja sitä pienentävät ostovelat (Knüpfer ja Puttonen 2018, s. 244.) Veroja ja käyttöpääoman muutoksia ei huomioida. Mallia omaisuuden arvon säilyttämiseksi perustellaan yksinkertaisuudella. Opinnäytetyössä pitäydytään liikevaihdon ja kustannusten käyttäytymisen perusasioiden tarkastelussa.

Kokonaisvarallisuuden ylläpitotarpeen mallintamiseksi malli sisältää teoreettisen korvausinvestointitarpeen arvioinnin, joka perustuu omaisuuserien taloudelliseen pitoaikaan. Olkoon T omaisuuden suunniteltu pitoaika vuosissa. Kun omaisuuskannan ikäjakauma oletetaan stationaariseksi eli tasaisesti jakautuneeksi välille $[0, T]$, jäljellä olevan pitoajan odotusarvo on $1/T$. Tällöin koko omaisuuskanta on teoreettisesti uusittava keskimääräisen jäljellä olevan pitoajan kuluessa. Teoreettinen vuotuinen saneeraustarve I_t^{teor} on

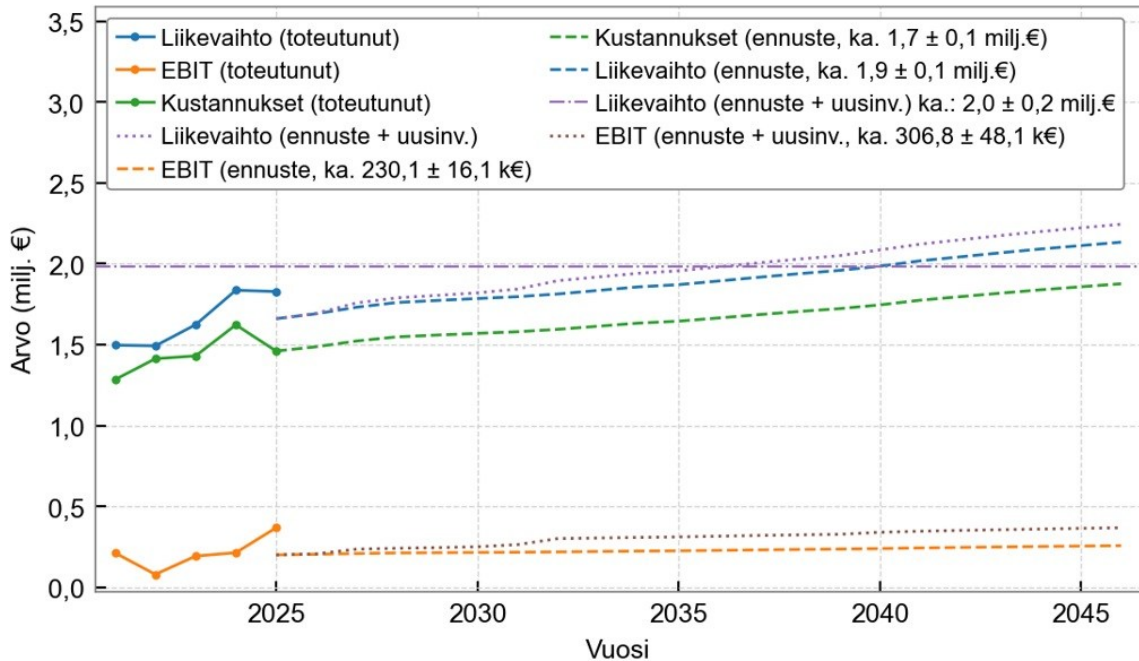
$$I_t^{\text{teor}} = T \times A_t. \quad (41)$$

Saneerausvaje eli korjausvelka vuonna t on toteutuvan saneerauksen ja teoreettisen tarpeen erotus

$$\Delta I_t = I_t^{\text{san}} - I_t^{\text{teor}}. \quad (42)$$

Negatiivinen ΔI kertoo korjausvelan kertymisestä, positiivinen ylijäämäisestä investointitasosta. Kumulatiivinen saneerausvaje saadaan summaamalla vuotuiset

vajeet tarkastelujakson yli. Negatiivinen kumulatiivinen vaje viittaa kasvavaan korjausvelkaan, joka on pitkällä aikavälillä yhteydessä vesihuoltolaitoksen operatiivisiin riskeihin. Kuvio 9 nähdään vesihuoltolaitoksen tarkasteluhetken liikevaihtotavoitteen vähimmäismäärä ajan funktiona, jotta omaisuuden laskennallinen arvo pysyisi vähintään samana.



Kuvio 9. Vähimmäisliikevaihto ja -voitto omaisuuden keskimääräisen arvon säilyttämiseksi case-vesihuoltolaitoksella.

3.3.6 Soveltaminen investointisuunnittelussa

Pitkän aikavälin suunnittelumallissa investoinnit kytketään DuPont-laskentaan erittelemällä niiden vaikutus omaisuuden arvoon, tuottotavoitteisiin ja kassavirtaan. Investoinnit jaetaan kahteen kategoriaan, joilla on erilaiset taloudelliset vaikutukset. Saneerausinvestoinnit ylläpitävät nykyistä omaisuuskantaa, eivätkä muuta taseen loppusummaa merkittävästi taseen reaaliarvoon nähden, koska ne korvaavat poistuvia omaisuuseriä. Niiden vaikutus on ensi sijassa kassavirtaperusteinen. Uusinvestoinnit kasvattavat laskennallista omaisuuskantaa ja siten taseen loppusummaa. Mallissa uusinvestointi I_t^{uusi} lisätään omaisuuden arvoon

$$A_t = A_{t-1} \times (1 + \pi_t) + I_t^{\text{uusi}}. \quad (43)$$

Uusinvestointi vaikuttaa sekä ROTA-yhtälön osoittajaan (tulosvaikutus) että nimittäjään (tasearvovaikutus). DuPont-laskennassa tämä näkyy siten, että uusinvestointi kasvattaa tarvittavaa liikevoittoa, koska $EBIT^{tavoite} = ROTA^{tavoite} \times A_t$ kasvaa, mutta samanaikaisesti se luo uutta tuottopotentiaalia, joka mahdollistaa suuremman liikevaihdon. Koska uusinvestoinnin nimellinen arvo sisältää investointihetken hankintahinnan, mutta tarkasteluhetken reaaliarvo laskee inflaation myötä, malli deflatoi kunkin uusinvestoinnin tarkasteluhetken kumulatiivisella inflaatiokertoimella

$$I_{i,t}^{\text{uusi, reaaliarvo}} = \frac{I_i^{\text{uusi}}}{\prod_{j=i+t}^t (1 + \pi_j)}. \quad (44)$$

Omaisuuuden kokonaisarvo A_t^{tot} uusinvestoinnit huomioiden on tällöin

$$A_t^{\text{tot}} = A_t + \sum_{i=0}^{t-1} I_{i,t}^{\text{uusi, reaaliarvo}}. \quad (45)$$

Tämä inflaatiokorjaus on merkityksellinen, koska se estää uusinvestointien ylivostuksen pitkällä aikavälillä. Esimerkiksi vuonna 2030 tehty yhden miljoonan euron uusinvestointi on reaaliarvoltaan vuonna 2045 huomattavasti pienempi, mikä vaikuttaa suoraan teoreettisen saneeraustarpeen ja tuottotavoitteiden laskentaan.

DuPont-järjestelmän avulla vesihuoltolaitoksen tavoitteiden realistisuutta voidaan arvioida kriittisesti. Liikevoittomarginaalin ja pääoman kiertonopeuden pitkän aikavälin kehitystä analysoimalla selvitetään, kykeneekö laitos vallitsevalla kustannusrakenteellaan ja investointitasollaan saavuttamaan vaaditun tuoton vai onko vesihuollon hinnoittelua korjattava. Vesihuoltolain 18 § ei suoraan määritä sallittua tuotto prosenttia, mutta edellyttää kustannusvastaavuutta, jonka mukaan maksutuloilla on katettava kaikki kustannukset pitkällä aikavälillä.

4 KONEOPPIMISEN PERIAATTEET

4.1 Python-ohjelmointikieli

Python on yksi suosituimmista ohjelmointikielistä tieteellisessä laskennassa, data-analytiikassa ja koneoppimisessa. Sen syntaksi on suunniteltu selkeäksi ja helposti omaksuttavaksi, mikä vähentää ohjelmoijan kognitiivista kuormaa ja mahdollistaa keskittymisen analyttisten ongelmien ratkaisemiseen teknisen toteutuksen sijaan (Sumathi ym., 2022, s. 15). Käännettävistä kielistä poiketen Python hyödyntää tulkkiarkkitehtuuria, jossa lähdekoodi muunnetaan ensin tavukoodiksi ja Python-virtuaalikone suorittaa tavukoodin rivi kerrallaan (Sumathi ym., 2022, s. 15–16). Tämä mahdollistaa välittömän palautteen syntaksi- ja suoritus-aikaisista virheistä, mikä nopeuttaa koodin testaamista ja kehittämistä.

Pythonin keskeinen asema koneoppimisessa perustuu sen laajaan ja nopeasti kehittyvään kirjastoekosysteemiin, joka kattaa data-analyysin keskeiset työvaiheet datan keruusta visualisointiin. Joshi ja Tiwari (2023) jäsentävät Python-kirjastoekosysteemin kolmeen toiminnalliseen kategoriaan: datan keruuseen, datan analysointiin ja prosessointiin sekä datan visualisointiin. Datan käsittelyssä keskeisimpiä kirjastoja ovat Pandas, joka tarjoaa joustavat tietorakenteet taulukkomuotoisen datan hallintaan, sekä NumPy, joka mahdollistaa tehokkaan monidimensionaalisten taulukoiden käsittelyn. Visualisoinnin puolella keskeisiä työkaluja ovat Matplotlib ja siihen pohjautuva Seaborn, jotka tukevat julkaisuihin soveltuvien graafisten esitysten tuottamista. (Joshi & Tiwari, 2023, s. 85.) Liitteessä 1 on esitetty koodiesimerkki, joka hyödyntää kaikkia kolmea kategoriaa: ecldata-kirjastoa datan hakemiseen Euroopan keskuspankin Data Portalista, Pandasia aikasarjan käsittelyyn ja Matplotlibia visualisointiin.

Sumathi ym. (2022, s. 15–17) korostavat, että Pythonin keskeisiä suunnitteluperiaatteita ovat luettavuus ja yksinkertaisuus. Tämä näkyy erityisesti pakollisessa sisennyksessä, joka toimii rakenteiden rajauksena ja korvaa monissa muissa kielissä käytetyt aaltosulkeet. Myös kontrollirakenteet eli ehtolauseet (engl. if, elif, else) ja toistorakenteet (engl. for, while) noudattavat suoraviivaista syntaksia, mikä tukee koodin luettavuutta. (Sumathi ym., 2022, s. 24–33.)

Data-analytiikassa Pythonia on tarkoituksenmukaista tarkastella osana laajempaa tietoputkea (engl. data pipeline), jossa raakadata etenee keruusta esikäsittelyyn, analyysiin, mallinnukseen ja raportointiin. Sumathi ym. (2022, s. 103–111) korostavat, että huomattava osa työmäärästä kohdistuu datan esikäsittelyyn, muotoiluun ja virheiden tunnistamiseen eikä varsinaiseen mallinnukseen. Pythonin virheenkäsittelymekanismien, erityisesti try-except-rakenteiden, merkitys korostuu, koska tietoaineistot sisältävät tyypillisesti puuttuvia arvoja ja rakennevirheitä, jotka on käsiteltävä robustisti tietoputken keskeytymättömän suorittamisen turvaamiseksi. (Sumathi ym., 2022, s. 74–75.)

Koneoppimisen ja tilastomallinnuksen näkökulmasta Python tarjoaa laaja-alaisen ja kerroksellisen ekosysteemin. Scikit-learn-kirjasto yhtenäistää koneoppimismallien käytön tarjoamalla yhteisen fit-predict-rajapinnan, mikä tukee mallien vertailtavuutta. Syväoppimisen alueella TensorFlow, Keras ja PyTorch muodostavat alan keskeisen työkalukentän. Niiden matriisilaskennan optimoinnit ja edistyneet rajapinnat mahdollistavat monimutkaisten neuroverkkorakenteiden toteutuksen. (Sumathi ym., 2022, s. 92–94; Joshi & Tiwari, 2023, s. 86–89.)

4.2 Koneoppiminen

Koneoppiminen (engl. machine learning, ML) on laskennallisen tietojenkäsittelyn osa-alue, jossa algoritmit oppivat datasta ilman sääntöpohjaista ohjelmointia ja parantavat suoriutumistaan kokemuksen karttuessa. Deuschlen (2019) mukaan koneoppiminen tarkoittaa tilastollisten, matemaattisten ja numeeristen menetelmien hyödyntämistä tiedon löytämiseksi ja johtamiseksi datasta. Tämä tieto voi ilmetä esimerkiksi tiivistämisenä, visualisointina, ryhmittelynä tai ennusteena. Keskeisenä lähtökohtana on kyky tunnistaa datasta säännönmukaisuuksia ja yleistää opitut riippuvuussuhteet uusiin tapauksiin. (Deuschle, 2019, s. 1–2.)

Kuviossa 10 esitetään koneoppimisen sykli, joka etenee datan keräämisestä esikäsittelyyn, analysointiin, mallin opettamiseen, testaamiseen ja hyödyntämiseen. Syklisyys korostaa prosessin iteratiivista luonnetta, koska mallia voidaan päivittää uuden datan ja arviointitulosten perusteella.



Kuvio 10. Koneoppimisen kuusivaiheinen sykli (Sumathi ym., 2022, s. 105, muokattu).

Koneoppimisessa erotetaan perinteisesti kaksi pääasiallista lähestymistapaa: ohjattu oppiminen ja ohjaamaton oppiminen (Deuschle, 2019, s. 2–3). Ohjatussa oppimisessa mallille annetaan opetusdataa, jossa jokaiseen syötteeseen liittyy ennalta tunnettu vaste. Tämä vaste voi olla esimerkiksi luokittelutehtävässä kohteen oikea luokka tai regressiotehtävässä jatkuva vastemuuttuja. Mallin tavoitteena on oppia ennustamaan vaste uusille havainnoille (Sonabend, 2021, s. 38). Ohjaamattomassa oppimisessä data ei sisällä ennalta annettuja vasteita, vaan mallin tehtävänä on löytää datasta piileviä rakenteita, kuten klustereita tai dimensioiden välisiä riippuvuuksia. (Deuschle, 2019, s. 2–3).

Ghosal (2023) korostaa, että koneoppimisen ytimessä on yleistettävyyden käsite: mallin tulee kyetä tuottamaan tarkkoja ennusteita myös sellaisista syötteistä, joita se ei ole opetusvaiheessa kohdannut. Yleistettävyyden saavuttaminen edellyttää huolellista mallinvalintaa ja parametrien säätöä, jotta vältetään sekä alioppiminen että ylioppiminen (Ghosal, 2023, s. 18–19). Deuschlen (2019, s. 21, 24) mukaan ylioppiminen ilmenee tilanteessa, jossa malli mukautuu opetusdataan liian tarkasti ja menettää kyvyn yleistyä uusiin havaintoihin, kun taas alioppimisessa malli ei reagoi riittävästi datan sisältämään vaihteluun.

Koneoppimisen algoritmit voidaan luokitella yksinkertaisiin ja monimutkaisiin malleihin niiden ilmaisuvoiman perusteella. Yksinkertaisiin malleihin lukeutuu lineaar-

rinen regressio, joka soveltuu tilanteisiin, joissa syötteen ja vastemuuttujan välinen suhde on lineaarinen (Deuschle, 2019, s. 6–7). Neuroverkot puolestaan muodostavat erittäin ilmaisuvoimaisen malliluokan, joka kykenee approksimoimaan mielivaltaisia funktioita riittävän monikerroksisella arkkitehtuurilla (Deuschle, 2019, s. 48–50; Berner, 2023, s. 15–16).

Koneoppimisen matemaattinen perusta rakentuu tilastotieteen, optimointiteorian ja lineaarialgebran varaan (Berner, 2023, s. 17–18; Deuschle, 2019, s. 8–10). Mallin oppiminen muotoillaan tyypillisesti optimointiongelmaksi, jossa tavoitteena on löytää sellaiset parametriarvot, jotka minimoivat valitun loss-funktion eli häviöfunktion (Deuschle, 2019, s. 8–10; Sonabend, 2021, s. 39). Esimerkiksi lineaarisessa regressiossa tavoitteena on löytää regressiokertoimet, jotka minimoivat ennusteiden ja havaittujen vasteiden välisen neliövirheen summan (Deuschle, 2019, s. 34).

Koneoppimisen soveltamiseen liittyy haasteita. Datan valmistelu vie koneoppimisprosessin vaiheista eniten aikaa, koska suuret aineistot on puhdistettava ja virheelliset tiedot korjattava huolellisesti. Lisäksi mallien tulkittavuus on keskeistä, sillä rajapintojen on oltava läpinäkyviä, jotta tulokset välittyvät ymmärrettävästi myös ilman datatieteen asiantuntemusta. Yksi koneoppimisen alalaji on syväoppiminen (engl. deep learning, DL), jossa hyödynnetään neuroverkkoja ihmisen hermojärjestelmän rakennetta jäljitellen. Algoritmin arkkitehtuuri määräytyy piilokerrosten lukumäärän mukaan, ja monikerroksiset rakenteet oppivat datasta piirteitä eri tarkkuuksilla ilman manuaalista piirresuunnittelua. (Sumathi ym., 2022, s. 4–7.)

Neuroverkot (engl. neural networks, NN) koostuvat kerroksittain järjestetyistä laskentayksiköistä, joissa jokainen yksikkö laskee syötteidensä painotetun summan ja soveltaa tulokseen epälineaarista aktivointifunktiota. Neuroverkon oppiminen tapahtuu säätämällä yhteyksien painoja siten, että verkon tuottamat ennusteet vastaavat mahdollisimman tarkasti opetusdatan vasteita (Deuschle, 2019, s. 48–58).

4.3 Syväoppiminen

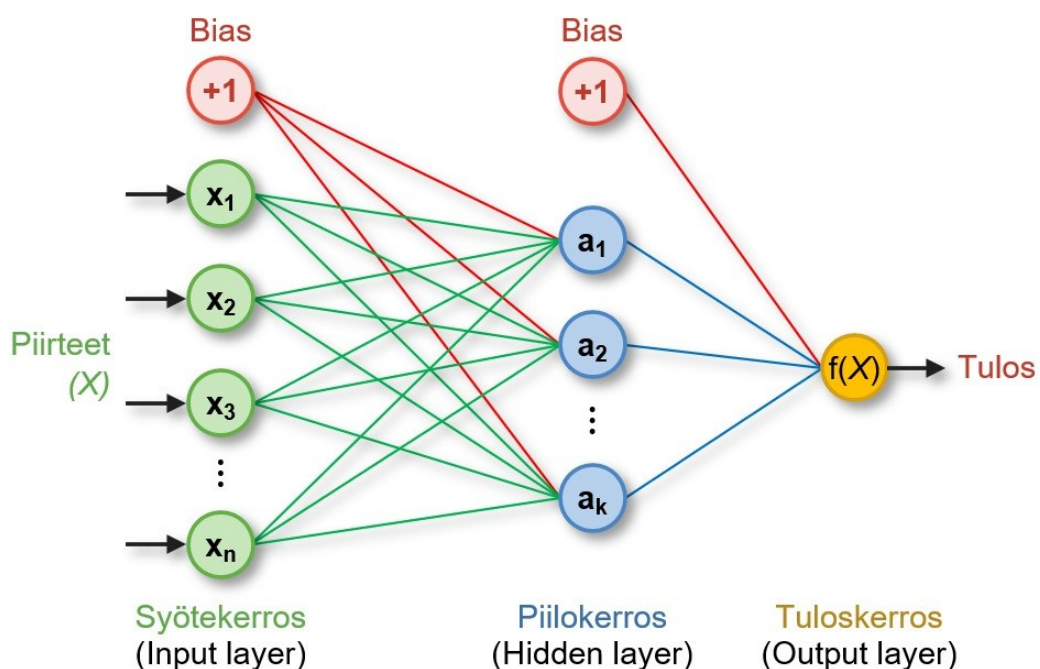
Syväoppiminen on koneoppimisen osa-alue, jossa monikerroksisia neuroverkkoja (engl. multi-layer perceptron, MLP) hyödynnetään monimutkaisten data-aineistojen analysointiin ja aineistosta opittavien piirteiden automaattiseen tunnistamiseen. Yksinkertaisin neuroverkon malli on perseptroni, joka koostuu yhdestä logistisesta yksiköstä (Julian, 2018, s. 55). Perseptronissa jokainen syöte kerrotaan omalla painokertoimellaan, ja näin saatu painotettu summa viedään aktiivointifunktioon, joka tuottaa verkon ulostulon. (Saleh, 2020, s. 28.) Perseptronin toiminta voidaan esittää matemaattisesti seuraavilla kahdella kaavalla

$$h_{\theta}(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n = \theta^T x, \quad (46)$$

$$h_{\theta}(x) = g \times (\theta^T x), \quad (47)$$

missä g on aktiivointifunktio, θ_n ovat painokertoimia ja x_n ovat syötteitä. Niin kutsuttu bias-syöte on $x_0 = 1$, jolloin painokerroin θ_0 toimii bias-terminä. (Julian, 2018, s. 49, 55.) Merkintä $\theta^T x$ tarkoittaa painokertoimien ja syötteiden pistetuloa, jossa yläindeksi T on transpoosi.

Kuviossa 11 on esitetty yhden piilokerroksen monikerroksinen perseptroni. Syötteet $x_1 \dots x_n$ kulkevat painotettujen yhteyksien kautta piilokerroksen neuroneille $a_1 \dots a_k$. Piilokerroksen ulostulot välittyvät edelleen tuloskerroksen funktiolle $f(X)$. Kuhunkin kerrokseen on lisätty erillinen bias-solmu.



Kuvio 11. Yhden piilotetun kerroksen MLP (scikit-learn, ei pvm., muokattu).

Täysin kytketyssä rakenteessa jokaisen kerroksen neuroni on yhdistetty kaikkiin seuraavan kerroksen neuroneihin. Verkon kerrokset jaetaan syötekerrokseen (engl. input layer), joka vastaanottaa datan, piilokerrokseen (engl. hidden layers), jotka käsittelevät datan, sekä tuloskerrokseen (engl. output layer), joka tuottaa lopputuloksen. Verkkoja, joissa on yli neljä kerrosta, pidetään tyypillisesti syvinä neuroverkkoina. (Saleh, 2020, s. 42.)

Aktivointifunktiot mahdollistavat epälineaarisuuden mallintamisen neuroverkossa. Ne ovat matemaattisia funktioita, jotka muuntavat neuronin painotetun syötteen lähtöarvoksi ja määräävät, kuinka voimakkaasti neuroni aktivoituu. Aktivointifunktion keskeinen tehtävä on rikkoa mallin lineaarisuus, mikä mahdollistaa monimutkaisten epälineaaristen yhteyksien oppimisen. (Saleh, 2020, s. 32). Erityisesti syvissä neuroverkoissa piilokerrosten yleisin aktivointifunktio on ReLU (engl. Rectified Linear Unit), sillä se on laskennallisesti tehokas ja toimii käytännössä hyvin (Julian, 2018, s. 72, 84). ReLU määritellään kaavalla

$$\text{ReLu}(x) = \max(0, x), \quad (48)$$

eli funktio palauttaa negatiivisille syötteille nollan ja säilyttää positiiviset arvot muuttumattomina.

Neuroverkon koulutusprosessi etenee vaiheittain. Ensin syötedata kulkee verkon läpi kerros kerrokselta niin sanotussa eteenpäin suuntautuvassa propagaatiossa (engl. forward propagation), jossa tieto siirtyy syötekerroksesta tuloskerrokseen. Jokainen neuroni käsittelee syötteensä painojen, biasin ja aktivointifunktion avulla ja välittää tuloksensa eteenpäin, kunnes viimeinen kerros saavutetaan ja verkko tuottaa ennusteen. (Saleh, 2020, s. 32.) Ennustetta verrataan todelliseen arvoon häviöfunktion avulla, joka mittaa numeerisesti ennusteen ja todellisen arvon välistä virhettä. Regressiot tehtävissä yleisin häviöfunktio on keskineliövirhe (engl. Mean Squared Error, MSE), joka määritellään seuraavasti (Saleh, 2020, s. 37.)

$$\text{loss}_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (49)$$

missä n on näytteiden lukumäärä, y_i on todellinen arvo ja \hat{y}_i on mallin ennuste. Häviöfunktion tavoitearvo on nolla, mikä tarkoittaa, ettei ennusteen ja todellisen arvon välillä ole lainkaan poikkeamaa. (Saleh, 2020, s. 36).

Häviön laskemisen jälkeen suoritetaan taaksepäin suuntautuva propagaatio (engl. backpropagation), jossa häviöfunktion osittaisderivaatat eli gradientit lasketaan verkon läpi takaisinpäin (Saleh, 2020, s. 37). Gradientit kertovat, kuinka paljon häviö muuttuu kunkin painon tai bias-termin pienen muutoksen seurauksena. Näiden perusteella mallin parametreja päivitetään gradienttimenetelmällä. Parametrien päivityssääntö voidaan määrittellä muodossa

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1), \quad (50)$$

missä α on oppimisnopeus (engl. learning rate) ja $\partial J(\theta)/\partial \theta_j$ on häviöfunktion osittaisderivaatta parametrin θ_j suhteen (Julian, 2018, s. 47). Käytännössä parametreja siirretään gradientin vastakkaiseen suuntaan, ja oppimisnopeus säätelee askelpituutta (Saleh, 2020, s. 39). Erägradienttimenetelmässä (engl. batch gradient descent) gradientti lasketaan koko koulutusaineiston yli jokaisella päivityskerralla, mikä on suurilla aineistoilla laskennallisesti raskasta. Stokastinen gradienttimenetelmä (engl. stochastic gradient descent, SGD) ratkaisee ongelman käsittelemällä yhden näytteen tai pienen satunnaisesti valitun erän kerrallaan, mikä nopeuttaa koulutusta ja mahdollistaa suurten aineistojen käytön. (Julian, 2018, s. 48; Saleh, 2020, s. 40.)

Neuroverkot oppivat toistamalla eteenpäin- ja taaksepäin-propagaatiota useita kertoja (Saleh, 2020, s. 31). Yhdellä epookilla (engl. epoch) tarkoitetaan yhtä täyttä läpikäyntiä koko koulutusaineiston yli (Julian, 2018, s. 66). Kunkin epookin aikana verkko suorittaa eteenpäin suuntautuvan propagaation, häviön laskennan ja taaksepäin suuntautuvan propagaation, joten koulutusprosessi on luonteeltaan iteratiivinen (Saleh, 2020, s. 31). Jokaisen epookin myötä malli tarkentuu. Mutta käytännössä malli ei saavuta täydellistä tarkkuutta. Koulutus päätetään, kun suorituskyky on riittävä tai kun malli ei enää parane.

Mallit suorittavat tyypillisesti useita satoja tai tuhansia epookkeja, ja koulutuksen onnistumista voidaan arvioida seuraamalla häviöfunktion arvon kehitystä epookkien edetessä. Arvon systemaattinen pieneneminen osoittaa, että malli oppii koulutusaineiston rakenteita ja kykenee mallintamaan syötemuuttujien ja kohde muuttujan välisen yhteyden. Epookkien lukumäärä on keskeinen hyperparametri, jonka optimaalinen arvo riippuu aineiston koosta, mallin monimutkaisuudesta sekä käytetystä oppimisnopeudesta.

4.4 Syväoppimisen kehitysympäristö

4.4.1 PyTorch-syväoppimiskehys

PyTorch on avoimen lähdekoodin syväoppimisen kirjasto, joka pohjautuu Torch-kirjastoon. Torch on alun perin toteutettu C-kielellä, ja PyTorch kehitettiin tuomaan sen tarjoamat toiminnot helposti käytettäväksi Python-ympäristössä. PyTorchin perusdatarakenne on tensori, joka on moniulotteinen matriisirakenne numeerisen datan esittämiseen. Tensorit muistuttavat rakenteeltaan NumPy-kirjaston taulukoita, mutta niiden keskeinen etu on siirrettävyys grafiikkaprosessorille (GPU), mikä mahdollistaa numeeristen laskentatoimitusten merkittävän nopeuttamisen (Saleh, 2020, s. 7). Automaattisen differoinnin toteuttamiseksi PyTorch hyödyntää autograd-järjestelmää, joka laskee gradientit tensoreihin kohdistuvista operaatioista (Julian, 2018, s. 61–62).

PyTorchin nn-moduuli tarjoaa korkean tason komponentteja neuroverkkojen toteuttamiseen. Yksinkertaisimmillaan verkon arkkitehtuuri määritellään nn.Sequential-luokalla, johon kerrokset lisätään peräkkäin halutussa järjestyksessä (Saleh, 2020, s. 14). Monimutkaisemmat verkot määritellään sen sijaan omaksi luokakseen, joka periytyy nn.Module-kantaluokasta. Periytymisen ansiosta malli saa käyttöönsä kaikki nn.Module-luokkaan kapseloidut toiminnallisuudet (Julian, 2018, s. 65).

Mallin koulutusyksi noudattaa kolmivaiheista rakennetta. Ensin suoritetaan eteenpäin suuntautuva propagaatio, jossa malli tuottaa ennusteen koulutusaineiston perusteella. Tämän jälkeen lasketaan häviö, ja taaksepäin suuntautuva propagaatio tuottaa häviöfunktion gradientit. Lopuksi mallin parametrit päivitetään gradienttien perusteella (Julian, 2018, s. 66; Saleh, 2020, s. 18). Gradientit on nollattava koulutussilmukan jokaisen iteraation alussa, jotta aiemmilta iteraatioilta kertyneet arvot eivät vääristä mallin oppimisprosessia (Julian, 2018, s. 67).

PyTorchin optim-paketti sisältää useita optimointialgoritmeja, kuten Adam-optimoijan, stokastisen gradienttilaskun (SGD) ja RMSprop-algoritmin (Saleh, 2020,

s. 17). Adam (engl. adaptive moment estimation) on optimointialgoritmi, jota käytetään neuroverkon painojen päivittämiseen koulutuksen aikana (Julian, 2018, s. 122). Tavallinen gradienttilasku käyttää samaa kiinteää oppimisnopeutta kaikille parametreille, kun taas Adam mukauttaa oppimisnopeuden jokaiselle paramet-rille erikseen (Julian, 2018, s. 115). Mukautus perustuu gradientin kahteen tilas-tolliseen tunnuslukuun: gradientin liukuvaan keskiarvoon ja gradientin neliöiden liukuvaan keskiarvoon, joita kutsutaan gradientin ensimmäiseksi ja toiseksi mo-mentiksi (Julian, 2018, s. 122).

4.4.2 PyTorch Lightning ja PyTorch Forecasting

PyTorch Lightning (Falcon & The PyTorch Lightning team, 2019) on PyTorchin päälle rakennettu avoimen lähdekoodin kirjasto, joka erottaa tutkimuksellisen lo-giikan toteutuksen rakennekoodista. Kirjasto automatisoi koulutussilmukan, lait-teiston hallinnan sekä hajautetun laskennan, jolloin ohjelmistokehittäjä voi keskit-tyä malliarkkitehtuuriin ja datan käsittelyyn. Arkkitehtuuri rakentuu kahden ydin-komponentin varaan. LightningModule kapseloi eteenpäin suuntautuvan propa-gaation, koulutus- ja validointivaiheet sekä optimoijan konfiguraation yhdeksi luo-kaksi, kun taas Trainer vastaa koulutuksen suorittamisesta ja skaalautuvuudesta. (Falcon & The PyTorch Lightning team, 2019.)

PyTorch Lightningin päälle rakennettu PyTorch Forecasting -kirjasto tarjoaa kor-kean tason rajapinnan aikasarjaennustamiseen. Sen keskeinen luokka TimeSe-riesDataSet toimii rajapintana Pandas DataFrame -datan ja PyTorch-mallien vä-lillä. Luokka edellyttää syöteaineistolta vähintään kolmea saraketta, joita ovat kasvava kokonaislukumuotoinen aikaleima, ennustettava kohdemuuttuja ja aika-sarjat yksilöivä ryhmätunniste. Luokka tukee myös staattisia ja ajassa muuttuvia kategorisia sekä jatkuvia muuttujia, mikä mahdollistaa monimuuttujaisen ennus-tamisen. (Beitner, ei pvm. a.)

Datan normalisoinnista vastaa target_normalizer-parametri, joka hyödyntää TorchNormalizer-, GroupNormalizer- tai EncoderNormalizer-luokkaa. Näistä En-coderNormalizer skaalaa datan dynaamisesti enkooderijakson perusteella, mikä estää mallia hyödyntämästä ennustehetkellä vielä tuntematonta tulevaisuuden

tietoa. TimeSeriesDataSet tukee lisäksi viivemuuttujia ja puuttuvien aikavaiheiden käsittelyä. Erillinen ennustetila tuottaa validointiin yhden ennustejakson kunkin aikasarjaa kohden. (Beitner, ei pvm. a.)

Mallin koulutus toteutetaan PyTorch Lightningin Trainer-luokan avulla. Tietoaineisto muunnetaan TimeSeriesDataSet-olioksi, jonka to_data_loader-metodilla muodostetaan koulutus- ja validointiaineiston lataajat. Trainer tukee EarlyStopping- ja LearningRateMonitor-takaisinkutsuja sekä TensorBoard-lokitusta. Oppimismnopeus voidaan etsiä automaattisesti Tuner-luokan lr_find-metodilla. PyTorch Forecasting tarjoaa lisäksi työkalut ennusteiden ja toteutuneiden arvojen vertailuun sekä muuttujien vaikutusten visualisointiin. Kirjaston keskeiset malliarkkitehtuurit on koottu taulukkoon 1. (Beitner, ei pvm. a.)

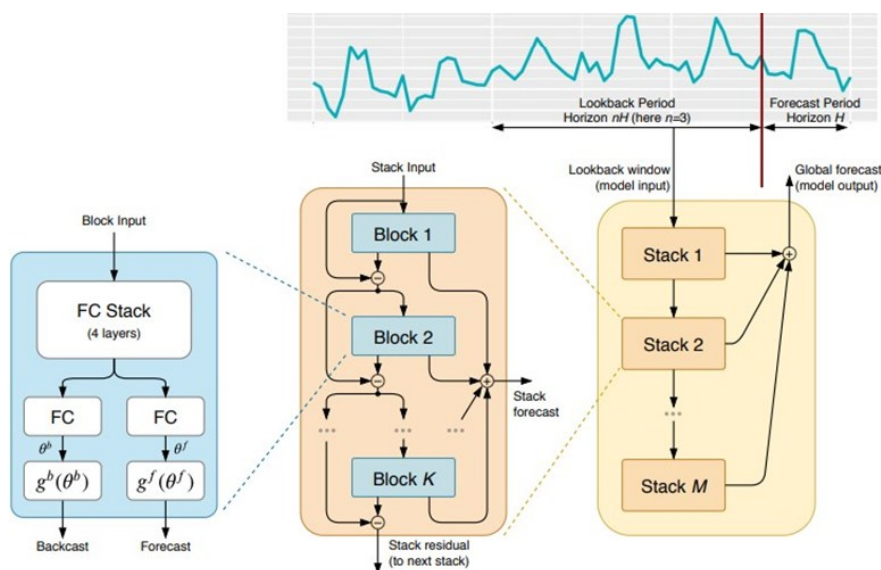
Taulukko 1. PyTorch Forecasting -mallivertailu (Beitner, ei pvm. a, muokattu).

Malli	Arkkitehtuuri	Kovariaatit	Monta kohdemuuttujaa	Epävarmuus	Laskentanopeus (1–5)	Huomiot
DecoderMLP	MLP-dekooderi	Kyllä	Kyllä	Ei	1	Ennuste ilman historiaa
DeepAR	Autoregressiivinen RNN	Kyllä	Kyllä	Kyllä	3	Hidas inferenssi
NBeats	Tulkittava/yleinen N-BEATS	Ei	Ei	Ei	1	Trendi ja kausi
NBeatsKAN	N-BEATS – KAN	Ei	Ei	Ei	1	Opitut spline-funktiot
NHiTS	Hierarkkinen interpolointi	Kyllä	Kyllä	Ei	2	Pitkät horisontit
RecurrentNetwork	LSTM/GRU-rekurrentti	Kyllä	Kyllä	Ei	2	Joustava historia
TemporalFusionTransformer	Huomiomekanismi – LSTM	Kyllä	Kyllä	Kyllä	4	Erinomainen tulkittavuus
TiDEModel	Enkooderi-dekooderi	Kyllä	Kyllä	Ei	3	Pitkän aikavälin ennusteet
xLSTM	Laajennettu LSTM (sLSTM/mLSTM)	Kyllä	Kyllä	Ei	3	Eksp. porttimek. – dekompositio

4.4.3 N-BEATS-neuroverkkoarkkitehtuuri

PyTorch Forecasting -kirjaston N-BEATS-luokka (engl. neural basis expansion analysis for interpretable time series forecasting, N-BEATS) on syväoppimiseen perustuva aikasarjaennustusmalli, jonka arkkitehtuuri rakentuu täysin kytkettyjen neuroverkkokerrosten (engl. fully connected layers, FC) varaan ilman aikasarja-analyysille tyypillisiä rekursiivisia tai konvoluutorakenteita. (Oreshkin ym., 2020, s. 2–3). Malli ei hyödynnä kovariaatteja, minkä vuoksi ainoan ennustajan on oltava kohdemuuttuja (Beitner, ei pvm. b). N-BEATS-arkkitehtuurin suunnittelua ohjaa kolme keskeistä periaatetta: Perusrakenteen on oltava yksinkertainen, yleiskäyttöinen ja silti ilmaisukykyinen. Arkkitehtuurin ei tule riippua aikasarjakoh- taisesta piirteiden muodostuksesta tai syötteen skaalauksesta. Sen on oltava laa- jennettavissa tulkittavaan muotoon. (Oreshkin ym., 2020, s. 2.)

Oreshkin ym. (2020) mukaan mallin perusrakenneyksikkö on haarautuva lohko (engl. block), joka koostuu nelikerroksisesta täysin kytketystä neuroverkosta sekä kahdesta lineaarisesta projektiosta (kuvio 12). Lineaariprojektiot tuottavat taak- sepäin- ja eteenpäinkertoimet θ^b ja θ^f , jotka syötetään kantalaajennuskerroksille. Näistä taaksepäinkerros g^b tuottaa takaisinennusteen (engl. backcast) \hat{x}_t , joka on lohkon arvio syöttestä, ja eteenpäinkerros g^f tuottaa ennusteen (engl. forecast) \hat{y}_t . Ensimmäisen lohkon syötteenä toimii kiinteän pituinen menneisyyden havain- tojen jakso eli lookback-ikkuna, kun taas seuraavien lohkojen syötteenä ovat edel- tävien lohkojen residuaaliulostuloja. (Oreshkin ym., 2020, s. 2–3.)



Kuvio 12. N-BEATS-arkkitehtuurin rakenne (Oreshkin ym., 2020, s. 3).

Lohkot on järjestetty pinoihin (engl. stacks). Pinojen ja lohkojen väliset residuaaliyhteydet mahdollistavat signaalin asteittaisen jalostamisen siten, että kukin lohko mallintaa edellisen lohkon jäännösvirhettä. Lopullinen ennuste muodostetaan summaamalla kaikkien lohkojen osittaisennusteet yhteen. (Oreshkin ym., 2020, s. 3–4; Sbrana & Lima de Castro, 2024, s. 1055.)

Oreshkin ym. (2020) määrittelevät N-BEATS-mallin yksimuuttujaisena pistemäisenä ennustusongelmana diskreetissä ajassa. Olkoon todellinen havaittu aikasarja $[y_1, \dots, y_T] \in \mathbb{R}^T$ pituudeltaan T . Tehtävänä on ennustaa tuleva vektori $y \in \mathbb{R}^H = [y_{T+1}, y_{T+2}, \dots, y_{T+H}]$, missä H on ennustehorisontti eli ennusteen askelmäärä. Mallin syötteenä käytetään lookback-ikkunaa $x \in \mathbb{R}^t = [y_{T-t+1}, \dots, y_T]$, missä $t \leq T$ on ikkunan pituus. Ikkunan pituus asetetaan tyypillisesti ennustehorisontin monikerraksi välille $2H-7H$. (Oreshkin ym., 2020, s. 2–3.)

Mallin ennustetarkkuuden arvioinnissa Oreshkin ym. (2020) hyödyntävät Hyndmanin ja Koehlerin (2006, s. 685) kehittämää MASE-mittaria (engl. mean absolute scaled error), joka normalisoi ennustevirheen suhteessa kausivaihtelun huomioivaan naiiviin vertailuennusteeseen. Mittari kehitettiin ratkaisuksi aiempien tarkkuusmittareiden, kuten MAPE:n, ongelmiin, jotka liittyvät nolllalla jakamisesta aiheutuviin määrittelemättömiin arvoihin ja mittareiden epäsymmetrisyyteen (Hyndman & Koehler, 2006, s. 683). MASE määritellään ennustehorisontin H yli seuraavasti

$$\text{MASE} = \frac{1}{H} \sum_{i=1}^H \frac{|y_{T+i} - \hat{y}_{T+i}|}{\frac{1}{T+H-m} \sum_{j=m+1}^{T+H} |y_j - y_{j-m}|}, \quad (51)$$

missä y_{T+i} on toteutunut arvo, \hat{y}_{T+i} sitä vastaavaa ennustetta ajanhetkellä $T+i$ ja m aikasarjan kausivaihtelun pituus. Osoittaja kuvaa absoluuttisten ennustevirheiden keskiarvoa ennustehorisontin yli. Nimittäjä muodostaa skaalaustekijän kausivaihtelun huomioivan naiivin ennustajan pohjalta. Naiivi ennustaja olettaa tulevan jakson arvon vastaavan jakson m aiemmin havaittua arvoa, jolloin nimittäjä tarjoaa luonnollisen referenssitason mallin virheen vertailuun. Arvo $\text{MASE} < 1$ osoittaa mallin tuottavan pienempiä virheitä kuin naiivi ennustaja, $\text{MASE} > 1$ kertoo päinvastaisesta ja $\text{MASE} = 1$ vastaa naiivin ennustajan tasoa. (Hyndman & Koehler, 2006, s. 685; Oreshkin ym., 2020, s. 2.)

Kunkin lohkon sisäinen laskenta tapahtuu kahdessa vaiheessa. Ensimmäisessä vaiheessa syöte kulkee neljän täysin kytketyn ReLU-kerroksen läpi, jotka tuottavat ketjutetusti piiloesitykset h_1 – h_4 . Viimeisestä piiloesityksestä h_4 johdetaan kahdella erillisellä lineaariprojektiolla eteenpäin- ja taaksepäinkertoimet θ^f ja θ^b (Oreshkin ym., 2020, s. 3.)

$$h_{l,1} = \text{FC}_{l,1}(x_l), \quad h_{l,2} = \text{FC}_{l,2}(h_{l,1}), \quad h_{l,3} = \text{FC}_{l,3}(h_{l,2}), \quad h_{l,4} = \text{FC}_{l,4}(h_{l,3}), \quad (52)$$

$$\theta_l^b = \text{Linear}_l^b(h_{l,4}), \quad \theta_l^f = \text{Linear}_l^f(h_{l,4}), \quad (53)$$

missä kukin FC-kerros on muotoa

$$h_{l,1} = \text{ReLU}(W_{l,1}x_l + b_{l,1}), \quad (54)$$

ja Linear on lineaarinen projektiokerros, eli

$$\theta_l^f = W_l^f h_{l,4}. \quad (55)$$

Kaava (52) osoittaa, että ekspansiokertoimet θ^b ja θ^f johdetaan samasta piirrektorista $h_{l,4}$ erillisillä lineaarisilla projektiokerroksilla. Oreshkinin ym. (2020, s. 3) mukaan eteenpäinekspansiokertoimien θ^f tehtävänä on optimoida osittaisennusteen \hat{y}_l tarkkuus sekoittamalla kantavektoreita, kun taas taaksepäinekspansiokertoimet θ^b tuottavat arvion syötteestä x_l tarkoituksenaan poistaa signaalista ne komponentit, jotka eivät edistä seuraavien lohkojen ennustamista.

Toisessa vaiheessa kantakerrokset θ^b ja θ^f projisoivat ekspansiokertoimet kantavektorien lineaarikombinaatioiksi tuottaen backcast- ja forecast-ulostulot muodossa

$$\hat{y}_l = \sum_{i=1}^{\dim(\theta_l^f)} \theta_{l,i}^f v_i^f, \quad \hat{x}_l = \sum_{i=1}^{\dim(\theta_l^b)} \theta_{l,i}^b v_i^b, \quad (56)$$

missä v_i^f ja v_i^b ovat forecast- ja backcast-kantavektorit ja $\theta_{l,i}^f$, i on vektorin θ_l^f i :s alkio. Kantakerrosten g^b ja g^f tehtävänä on tarjota riittävän rikkaat kantavektori-joukot $\{v_i^f\}$ ja $\{v_i^b\}$, joiden avulla ulostulot voidaan esittää ekspansiokertoimia θ^f ja θ^b vaihtelemalla. Kantavektorit voivat olla joko vapaasti opittavia tai ennalta määrättyjä funktiomuotoja, jotka tuovat malliin rakenteellisia oletuksia. (Oreshkin ym., 2020, s. 3–4.)

N-BEATS-arkkitehtuurin keskeinen ratkaisu on kaksinkertainen residuaalin silmukatopologia, joka kuvataan kaavoilla

$$x_l = x_{l-1} - \hat{x}_{l-1}, \quad \hat{y}_l = \sum_l \hat{y}_l, \quad (57)$$

missä ensimmäinen osa kuvaa backcast-residuaalihaaraa. Kunkin lohkon syötteestä x_l poistetaan edellisen lohkon backcast-ulostulo \hat{x}_{l-1} , jolloin edeltävä lohko poistaa signaalista sen osan, jonka se pystyy mallintamaan. Tämä helpottaa seuraavien lohkojen ennustetehtävää progressiivisesti. Kaavan (57) jälkimmäinen osa kuvaa forecast-koontia. Lopullinen ennuste \hat{y} on kaikkien lohkojen osittaisennusteiden \hat{y}_l summa. Hierarkkinen dekomponointi yhdistetään ensin pinotasolla ja sitten koko verkon tasolla. (Oreshkin ym., 2020, s. 4.)

Oreshkin ym. (2020) toteavat, että kaksinkertaisessa residuaalin silmukatopologiassa backcast-haara mahdollistaa syötesignaalin sekventiaalisen analyysin, kun taas osittaisennusteiden hierarkkinen yhdistäminen forecast-haarassa tuottaa tulkittavan dekomponoinnin. Kun kantakerrosten g^b ja g^f rakenne on määrätty ja jaettu pinon sisällä, osittaisennusteet voidaan summata merkityksellisistä komponenteista koostuvaksi kokonaisennusteeksi. (Oreshkin ym., 2020, s. 4.)

N-BEATS-arkkitehtuurista on kaksi konfiguraatiota, jotka määräytyvät kantakerrosten g^b ja g^f valinnan perusteella: yleinen (engl. generic) ja tulkittava (engl. interpretable). Yleinen konfiguraatio ei nojaa aikasarjakohtaiseen tietoon. Kantakerrokset g^b ja g^f asetetaan lineaarisiksi projektioiksi edellisen kerroksen ulostulosta, jolloin lohkon ulostulot ovat muotoa

$$\hat{y}_l = V_l^f \theta_l^f + b_l^f, \quad \hat{x}_l = V_l^b \theta_l^b + b_l^b, \quad (58)$$

missä V_l^f on $H \times \dim(\theta_l^f)$ -ulotteinen matriisi, jonka ensimmäinen dimensio tulkitaan diskreettiseksi aikaindeksiksi ennustealueella ja toinen dimensio kantafunktioiden indekseiksi. Ekspansiokertoimet θ_l^f toimivat näiden kantafunktioiden painokertoimina, joten matriisin V_l^f sarakkeita voidaan tulkita aaltomuotoina aika-alueella. Koska matriisiin V_l^f ei kohdisteta lisärajoitteita, neuroverkon oppimat aaltomuodot eivät ole rakenteellisesti tulkittavia. (Oreshkin ym., 2020, s. 4.) Tulkittava konfiguraatio hyödyntää samaa yleistä arkkitehtuuria, mutta kantakerrosten rakenteeseen lisätään yleistävää muutosta pinotasolla. Oreshkin ym. (2020, s. 4) ehdottavat trendi- ja kausittaisuusdekomponoinnin suunnittelua suoraan malliin, mikä noudattaa aikasarjan hajotelman menetelmiä.

Tyypillinen trendin ominaisuus on, että se on useimmiten yksisuuntainen tai ainakin hitaasti muuttuva funktio. Tämän käyttäytymisen mallintamiseksi kantakerrokset g_{s,b_l} ja g_{s,f_l} rajoitetaan matala-asteiseksi polynomiksi

$$\hat{y}_{s,l} = \sum_{i=0}^p \theta_{s,l,i}^f t^i, \quad (59)$$

missä aikavektori t koostuu kokonaisluvuista $[0, 1, 2, \dots, H-2, H-1]^T$ nollassa $H-1$:een asti. Kun aikavektori t jaetaan H :lla, saadaan diskreetti aikahila, joka alkaa arvosta 0 ja päättyy arvoon $(H-1)/H$. Matriisimuodossa trendiennuste on

$$\hat{y}_{s,l}^{\text{tr}} = T\theta_{s,l}^f, \quad (60)$$

missä $\theta_{s,l}^f$ ovat yhtälön FC-verkon ennustamat polynomikertoimet ja T on $[1, t, t^2, \dots, t^p]$ on potenssien matriisi. Kun p on matala, esimerkiksi 2 tai 3, ennuste $\hat{y}_{s,l}^{\text{tr}}$ pakotetaan jäljittämään trendiä. (Oreshkin ym., 2020, s. 5.)

Kausittaisuuden tyypillinen ominaisuus on säännöllinen, syklinen ja toistuva vaihtelu. Tämän mallintamiseksi kantakerrokset g_{s,b_l} ja g_{s,f_l} rajoitetaan periodisten funktioiden joukkoon (Oreshkin ym., 2020, s. 5). Periodisten funktioiden kanta-joukoksi teoreettisesti perusteltu valinta on Fourier-sarja, jonka avulla periodinen käyttäytyminen voidaan esittää sini- ja kosinifunktioiden avulla. Rakenne mahdollistaa mallin kyvyn kuvata säännöllisesti toistuvia vaihteluita eri taajuuksilla

$$\hat{y}_{s,l} = \sum_{i=0}^{\lfloor H/2-1 \rfloor} \theta_{s,l,i}^f \cos(2\pi it) + \theta_{s,l,i+\lfloor H/2 \rfloor}^f \sin(2\pi it). \quad (61)$$

Matriisimuodossa kausittaisuusennuste on

$$\hat{y}_{s,l}^{\text{seas}} = S\theta_{s,l}^f, \quad (62)$$

missä $\theta_{s,l}^f$ ovat Fourier-kertoimia, jotka ennustavat pinon s kerroksen l FC-verkossa. S on $[1, \cos(2\pi t), \dots, \cos(2\pi \lfloor H/2 - 1 \rfloor t), \sin(2\pi t), \dots, \sin(2\pi \lfloor H/2 - 1 \rfloor t)]$ on sinimuotoisten aaltomuotojen matriisi. Ennuste $\hat{y}_{s,l}^{\text{seas}}$ on siten periodinen funktio, joka jäljittelee tyypillisiä kausittaisia kuvioita. Kokonaisuutena tulkittava arkkitehtuuri koostuu kahdesta pinosta. Trendipino edeltää kausittaisuuspinoa. Kaksinkertainen residuaalipinoutuminen yhdistettynä backcast-forecast-periaatteeseen johtaa siihen, että trendikomponentti poistetaan syöteikkunasta x ennen sen syöttämistä kausittaisuuspinoon. Trendi- ja kausittaisuusennusteet ovat saatavilla erillisinä tulkittavina ulostuloina. (Oreshkin ym., 2020, s. 5.)

5 KONEOPETETTU OMAISUUDENHALLINTAMALLI

5.1 Artefaktin kehittäminen

Vesihuoltolain uudistus edellyttää vesihuoltolaitoksilta systemaattista omaisuudenhallintasuunnitelmaa, joka turvaa verkostojen pitkäjänteisen ylläpidon, investointien suunnittelun ja taloudellisen kestävyuden. Suunnitelman laadinnassa voidaan hyödyntää tarkoitukseen soveltuvia tietojärjestelmiä, joihin osa tiedoista voi sijoittua suoraan (HE 40/2025 vp, s. 85, 169–170).

Digitalisaatio nähdään strategisena edellytyksenä vesihuollon taloudelliselle ja toiminnalliselle kestävyydelle. Omaisuudenhallintasuunnitelman tulee toimia yhteenvedona yksityiskohtaisemmista suunnitelmista ja tukea läpinäkyvää omaisuudenhallintaa (HE 40/2025 vp, s. 85, 99), jolloin digitaaliset työkalut nousevat keskeiseksi keinoksi vastata alan haasteisiin.

Koneoppimiseen perustuvan omaisuudenhallintamallin kehittäminen vastaa tarpeeseen siirtyä manuaalisesta suunnittelusta dynaamiseen ja datalähtöiseen päätöksentekoon. Nykyiset menetelmät eivät riittävästi yhdistä investointi- ja taloussuunnittelua, joten tarvitaan lähestymistapa, jossa koneoppimisalgoritmit, tietojen integrointi ja visuaalinen analytiikka muodostavat yhtenäisen kehyksen pitkän aikavälin resurssien kohdentamiseen.

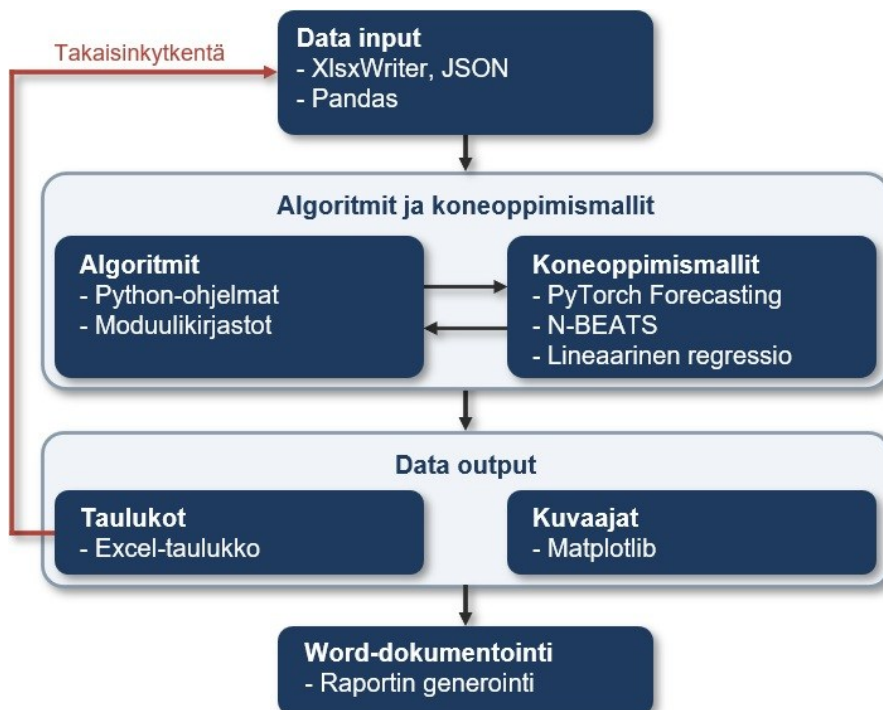
Malli toteutetaan integroidulla Python-ohjelmistoalustalla, joka kokoaa omaisuudenhallinnan tiedot yhteen ja mahdollistaa prosessien standardisoinnin sekä päätöksenteon läpinäkyvyyden. Koneoppiminen luo edellytykset mukautuvalle järjestelmälle, joka reagoi joustavasti toimintaympäristön muutoksiin.

Soveltaminen edellyttää kuitenkin systemaattista kehittämistä ja validointia: mallin luotettavuus riippuu aineiston laajuudesta ja laadusta, ja selitettävyys on julkisella sektorilla välttämätöntä perusteluvollisuuden ja sidosryhmien ymmärryksen vuoksi. Kokonaisuudessaan kyseessä on monitieteinen haaste, joka yhdistää teknisiä, hallinnollisia ja yhteiskunnallisia näkökulmia.

5.2 Arkkitehtuuri

Koneoppimiseen perustuvan omaisuudenhallintamallin arkkitehtuuri pohjautuu modulaariseen järjestelmäarkkitehtuuriin, joka yhdistää tiedon, koneoppimisalgoritmit, kustannusanalytiikan ja visuaalisen raportoinnin yhtenäiseksi analyytiseksi kokonaisuudeksi. Modulaarinen rakenne mahdollistaa eri laskenta- ja enustemoduulien joustavan yhdistämisen järjestelmäkokonaisuuksiin ja taloudellisiin taustatekijöihin, mikä tukee vesihuoltolaitoksen kokonaisvaltaista analyysiä. Lisäksi modulaarisuus helpottaa järjestelmän skaalautuvuutta, ylläpitoa, päivittämistä ja laajentamista tulevia tarpeita varten.

Kuviossa 13 esitetty arkkitehtuuri kuvaa kerroksittaista tiedonkäsittelykokonaisuutta, joka etenee tiedonkeruusta automatisoituun raportointiin. Ensimmäisessä kerroksessa rakenteinen lähtödata tuodaan järjestelmään Excel- ja JSON-muodoissa, ja sen esikäsittely sekä yhdistäminen toteutetaan Python-kirjastoja hyödyntäen. Tämä vaihe luo yhtenäisen lähtöaineiston ja varmistaa myöhempien analyysien toistettavuuden.



Kuvio 13. Omaisuudenhallintamallin ohjelmalliset moduulit.

Toisessa kerroksessa aineistoa käsitellään algoritmeista ja koneoppimismalleista muodostuvassa kokonaisuudessa. Algoritmiosio koostuu Python-ohjelmasta ja

sen käyttämistä moduulikirjastoista, kun taas koneoppimisosio hyödyntää PyTorch Forecasting -kirjastoon perustuvaa N-BEATS-mallia sekä lineaarista regressiota. Osien välinen kaksisuuntainen vuorovaikutus mahdollistaa algoritmisen esikäsittelyn ja mallien tuottamien tulosten yhdistämisen, jolloin sääntöpohjainen laskenta ja oppiva ennustaminen täydentävät toisiaan.

Kolmannessa kerroksessa tulokset tallennetaan Excel-taulukoiksi ja visualisoidaan Matplotlib-kirjastolla. Taulukkomuotoinen tuloste palautetaan takaisinkytkentänä tiedonkeruukerrokseen, jolloin laskennan tuloksia voidaan hyödyntää syötteinä jatkoajoissa. Lopuksi taulukot ja kuvaajat kootaan automaattisesti Word-dokumentiksi, joka muodostaa prosessin lopullisen tuotoksen – omaisuudenhallintasuunnitelman.

5.2.1 Tiedonkeruu ja yhdistäminen

Tiedonkeruu ja yhdistäminen muodostavat koneoppimispohjaisen omaisuudenhallintamallin perustan, koska analyysin luotettavuus riippuu lähtödatan kattavuudesta ja laadusta. Monilähteinen tietoinfrastrukturi yhdistää taloudelliset ja tekniset indikaattorit yhtenäiseksi kokonaisuudeksi, johon tuodaan kirjanpitojärjestelmistä tilinpäätös-, kustannus- ja kassavirtatiedot.

Datalähtöinen päätöksenteko edellyttää vesihuoltolaitoksilta analytiikkaosaamista ja valmiutta rajapintapohjaiseen tiedonsiirtoon. Ohjelmointirajapintojen (API) kautta toteutetut integraatiot mahdollistavat tietojen reaaliaikaisen päivittämisen, mikä tukee dynaamista ennustamista sekä skenaariopohjaista investointi- ja taloussuunnittelua. Omaisuudenhallintamalli voi näin reagoida toimintaympäristön muutoksiin ilman manuaalista päivitystyötä.

Integraation onnistuminen edellyttää tiedon laadunvarmistusta ja yhteensopivuutta lähdejärjestelmien välillä. Esikäsittelyssä tunnistetaan puuttuvat arvot, poikkeamat ja yksikkömuunnoksia vaativat kentät, sillä koneoppimisalgoritmien ennustetarkkuus määräytyy syötteiden laadun perusteella. Tiedonkeruu ja integraatio määrittävät viime kädessä järjestelmän analyttisen kapasiteetin ja kyvyn tuottaa päätöksenteon kannalta käyttökelpoista tietoa.

5.2.2 Koneoppimismoduulit

Koneoppimismoduulit muodostavat omaisuudenhallintamallin analyttisen ytimen, jossa ennustava analytiikka perustuu kahden toisiaan täydentävän menetelmän yhdistelmään teknisten ja taloudellisten muuttujien mallintamiseksi. Menetelmien yhdistäminen tarjoaa sekä syväoppimisen joustavuuden että tilastollisen tulkittavuuden, mikä vahvistaa ennusteiden luotettavuutta ja käyttökelpoisuutta päätöksenteon tukena.

Pääasiallisena ennustemallina toimii N-BEATS-arkkitehtuuri, joka ei edellytä oletuksia aikasarjan rakenteesta, vaan oppii sen olennaiset komponentit suoraan historiallisista havainnoista. Arkkitehtuurin modulaarinen lohkorakenne mahdollistaa ennusteen hajottamisen tulkittaviin osakomponentteihin, kuten trendiin ja kausivaihteluun, mikä parantaa mallin läpinäkyvyyttä ja lisää sen käyttökelpoisuutta päätöksenteon tukena.

Koneoppimista täydennetään lineaarisella regressioanalyysillä, joka mallintaa muuttujien välisiä riippuvuuksia. Regressiosuoran parametrit estimoidaan pienimmän neliösumman menetelmällä, joka minimoi havaittujen ja ennustettujen arvojen poikkeamien neliösumman. Mallin selitysvoimaa arvioidaan selitysteella R^2 . Vaikka menetelmä on yksinkertainen, se tarjoaa laskennallisesti tehokkaan ja tilastollisesti tulkittavan kehyksen taloudellisten muuttujien välisten lineaaristen riippuvuuksien tarkasteluun.

Järjestelmän toteutus perustuu Python-ohjelmointiympäristöön, joka mahdollistaa mallien automaattisen koulutuksen, validoinnin ja parametrien optimoinnin. Pythonin tarjoama automatisointi tukee mukautuvaa ennustamista ja reaaliaikaista analytiikkaa, jossa sekä N-BEATS-mallin tuottamat syväoppimiseen perustuvat ennusteet että regressioanalyysin kuvaamat muuttujien väliset riippuvuudet päivittyvät uuden datan myötä. Näin koneoppimismoduulit muodostavat perustan dynaamiselle omaisuudenhallintajärjestelmälle, joka tehostaa investointipäätösten pitkän aikavälin vaikutusten arviointia ja tukee riskienhallintaa kestäväen kehityksen periaatteiden mukaisesti.

5.2.3 Visualisoinnin merkitys

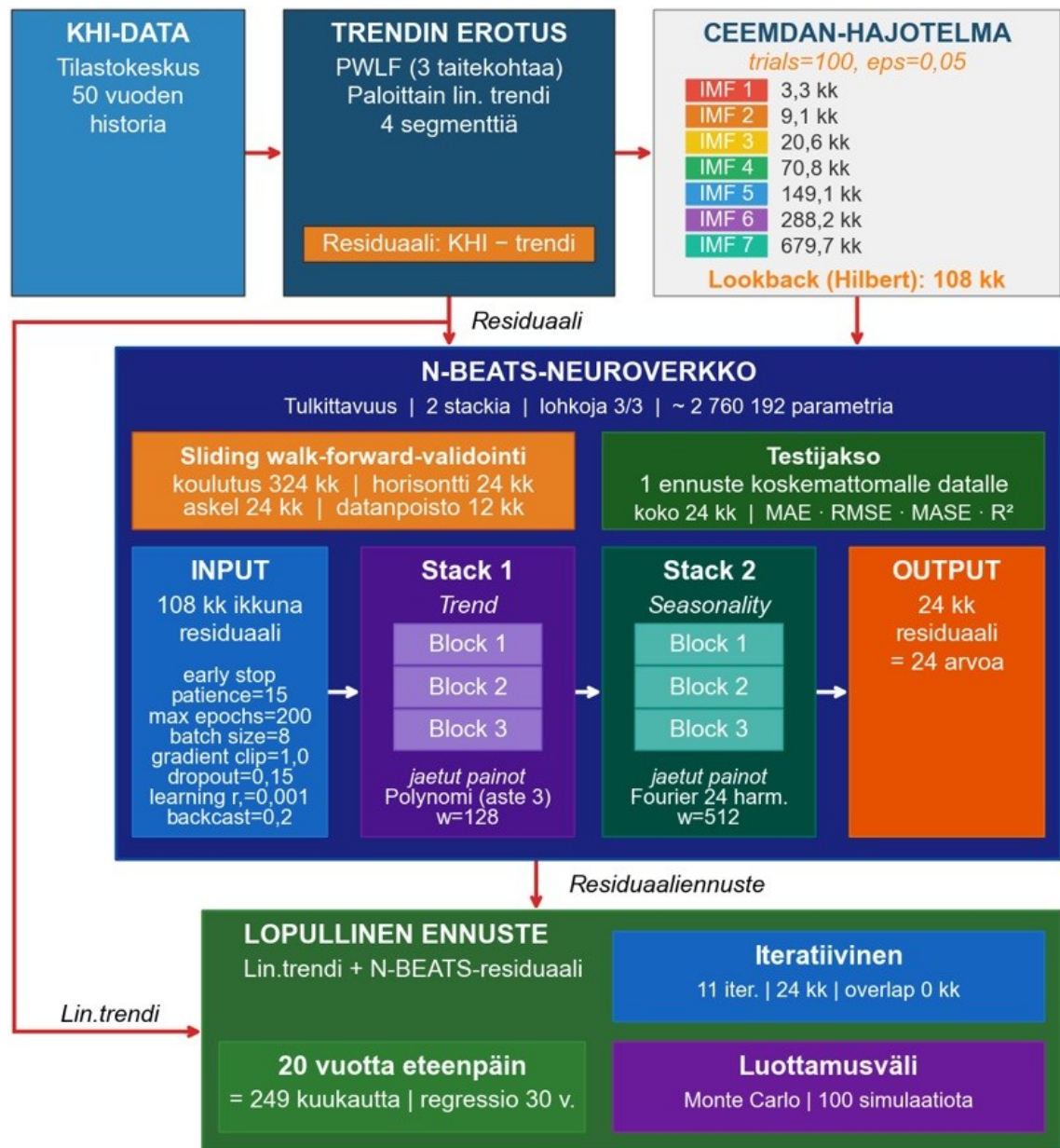
Visualisointi on omaisuudenhallintamallin keskeinen toiminnallisuus, jossa määrälliset analyysitulokset muunnetaan helpommin tulkittavaan muotoon. Automatisoitu prosessi tuottaa graafisia esityksiä ja taulukoita, jotka parantavat tiedon saavutettavuutta eri sidosryhmissä teknisestä taustasta riippumatta. Visuaalinen esitystapa välittää aineiston keskeiset säännönmukaisuudet tehokkaammin kuin pelkät taulukkomuotoiset arvot.

Graafinen esitystapa havainnollistaa taloudellisten tunnuslukujen ajallista kehitystä ja niihin liittyviä epävarmuustekijöitä, mikä tukee sekä nykytilanteen arviointia että tulevaisuuden skenaarioiden tarkastelua. Se on erityisen hyödyllinen pitkän aikavälin vertailussa, sillä vaihtoehtoiset kehityspolut voidaan esittää rinnakkain ja yhteismitallisesti.

Visualisointi toimii myös aineiston laadun arvioinnin välineenä, koska poikkeamat ja epä johdonmukaisuudet on usein helpompi tunnistaa graafisesta esityksestä kuin numeerisesta aineistosta. Esitystapaan liittyy kuitenkin rajoituksia. Se on numeerista aineistoa epätarkempi. Lisäksi huolimaton tai valikoiva esitystapa voi vääristää tulkintaa, joko tahattoman virheen tai tarkoituksellisen harhautuksen seurauksena. Tämän vuoksi graafisten esitysten laatiminen ja tulkinta edellyttävät kriittisyyttä ja menetelmällistä asiantuntemusta. Riski korostuu automatisoiduissa järjestelmissä, sillä esitysten vaivaton tuottaminen voi johtaa menetelmien käyttöön ilman riittävää ymmärrystä niiden taustaoletuksista.

5.3 Inflaatioanalyysi ja pitkän aikavälin inflaatioennuste

Omaisuudenhallintamallin ensimmäisessä vaiheessa suoritetaan inflaatioanalyysi syväoppimiseen perustuvalla ennustesovelluksella. Mallin arkkitehtuuri muodostaa prosessiketjun, jossa kuluttajahintaindeksin aikasarja käsitellään peräkkäisten menetelmävaiheiden kautta tulkittavaan muotoon. Aineisto haetaan Tilastokeskuksen StatFin PxWeb API -rajapinnasta viideltä viimeisimmältä vuosikymmeneltä. Kuviossa 14 esitetään PWLF–CEEMDAN–N-BEATS-mallin kokonaisarkkitehtuuri.



Kuvio 14. PWLF-CEEMDAN-N-BEATS-mallin tulkittava arkkitehtuuri KHI-datasta lopulliseen ennusteeseen.

Esikäsittelyssä KHI-aikasarjasta erotetaan pitkän aikavälin trendi paloittaisella lineaarisella regressiolla (engl. piecewise linear fitting, PWLF). Neljän segmentin ja kolmen taitekohdan avulla muodostetaan residuaali, joka kuvaa aikasarjan satunnaista ja kausittaista vaihtelua ilman hitaasti muuttuvaa tasokomponenttia. Trendin erotus estää sen, että historiallisesti merkittävät inflaatiopiikit tulkittaisiin pysyviksi kehityskuluiksi ja johtaisivat epärealistisiin ennusteisiin. Neljä segmenttiä riittää tunnistamaan 50 vuoden datan keskeiset käännekohtat: 1970-luvun öljykriisit, 1980-luvun lopun korkeasuhdanteen huipun, vuoden 2008 finanssikriisin

sekä vuoden 2020 COVID-shokin, jota seurasi vuosien 2021–2023 korkean inflaation kausi.

Residuaalisarja hajotetaan CEEMDAN-algoritmillä, joka on täydellinen joukkoistettu empiirinen moodinpurku mukautuvalla kohinalla (engl. complete ensemble empirical mode decomposition with adaptive noise). Algoritmi purkaa residuaalisygnaalin luontaisiin moodifunktioihin (engl. intrinsic mode function, IMF), joista kukin kuvaa erillistä taajuus- ja amplitudidynamiikkaa. IMF:lle arvioidaan dominoiva jakso Hilbert-muunnoksella, minkä pohjalta johdetaan suositus N-BEATS-verkon lookback-ikkunan pituudelle.

Esikäsitelty residuaali syötetään N-BEATS-neuroverkkoon, joka koostuu kahdesta pinosta: ensimmäinen mallintaa trendiä kolmannen asteen polynomilla ja toinen kausivaihtelua Fourier-sarjalla. Kumpikin pino sisältää kolme peräkkäistä lohkoa, joiden painot on jaettu pinon sisällä. Verkko vastaanottaa Hilbert-analyysin suosittelman konteksti-ikkunan ja tuottaa 24 kuukauden residuaaliennusteen. Koulutuksessa käytetään walk-forward-menettelyä, jossa 324 kuukauden koulutusikkuna liukuu 24 kuukauden askelin eteenpäin. Mallin suorituskyky arvioidaan erillisellä 24 kuukauden testijaksolla.

Lopullinen ennuste muodostetaan summaamalla 30 vuoden referenssijaksosta estimoitu lineaarinen trendi ja N-BEATS-verkon residuaaliennuste. Ennustehorisontti ulottuu kuluvan vuoden lopusta 20 vuotta eteenpäin, ja se tuotetaan iteratiivisesti ketjuttamalla 11 peräkkäistä 24 kuukauden ennusteaskelta.

Ennusteen epävarmuus kuvataan 50, 80 ja 95 prosentin luottamuskäyriä muodostavana viuhkakaaviona, joka lasketaan Monte Carlo -simulaatiolla. Simulaatiossa muunnellaan samanaikaisesti neljää riippumatonta epävarmuuden lähdettä: walk-forward-virheiden empiiristä jakaumaa, trendin kulmakertoimen ajassa lineaarisesti kasvavaa keskivirhettä, ajan neliöjuurena kasvavaa historiallista residuaalivolatiliteettia sekä N-BEATS-verkon sisäistä parametripäivämuuttoa. Yhdistelmä erottaa toisistaan rakenteellisen trendin, mallin sovitusepävarmuuden ja satunnaisten makrosokkien kasvavan vaikutuksen pitkällä ennustehorisontilla.

5.3.1 Trendin estimointi paloittaisella lineaarisella regressiolla

Regressioanalyysi on kvantitatiivisen tutkimuksen perusmenetelmä, jolla tarkastellaan selittävien muuttujien yhteyttä selitettävään muuttujaan. Menetelmän etuna on, että useita selittäviä muuttujia voidaan tutkia samanaikaisesti, jolloin yksittäisen muuttujan yhteys selitettävään muuttujaan voidaan arvioida muiden tekijöiden vaikutus vakioituna. Menetelmä edellyttää, että selitettävä muuttuja on vähintään välimatka-asteikollinen. (Kaakinen & Ellonen, ei pvm.)

Yksinkertaisen lineaarisen regressioanalyysin perusmalli esitetään yhtälöllä

$$Y = a + bX + \varepsilon, \quad (63)$$

jossa Y on selitettävän muuttujan arvo, a vakiotekijä, b regressiokerroin, X selittävän muuttujan arvo ja ε virhetermi. Regressiokerroin b ilmaisee, kuinka paljon selitettävän muuttujan ennustettu arvo muuttuu, kun selittävä muuttuja kasvaa yhdellä yksiköllä. Kahden selittävän muuttujan tapauksessa kaava laajenee muotoon

$$\hat{Y} = a + b_1X_1 + b_2X_2, \quad (64)$$

jossa \hat{Y} on selitettävän muuttujan ennustettu arvo, a vakiotekijä, X_1 ja X_2 selittävät muuttujat sekä b_1 ja b_2 niiden regressiokertoimet. Mallin selitysvoimaa arvioidaan R^2 -luvulla, joka ilmaisee selitettävän muuttujan kokonaisvaihtelun osuuden, jonka malli pystyy selittämään. Malleja vertailtaessa suositellaan korjattua R^2 -lukua, koska se huomioi mallissa käytettyjen muuttujien lukumäärän. (Kaakinen & Ellonen, ei pvm.)

Regressiokertoimien tilastollista merkitsevyyttä arvioidaan p-arvon avulla. P-arvo ilmaisee todennäköisyyden sille, että vastaava yhteys havaittaisiin, jos muuttujien välillä ei todellisuudessa olisi yhteyttä. Mitä pienempi arvo on, sitä epätodennäköisempi satunnaisselityksen mahdollisuus on. Yleisesti käytetyn 95 %:n luottamustason mukaan p-arvoa, joka on pienempi kuin 0,05, pidetään tilastollisesti merkitsevänä. (Kaakinen & Ellonen, ei pvm.)

Residuaalivektori on sovitetun regressiomallin ja alkuperäisten havaintojen välinen erotus kaikissa datapisteissä. Jekel ja Venter (2019, s. 4) määrittelevät residuaalivektorin e muodossa

$$e = A\beta - y, \quad (65)$$

missä A on regressiomatriisi, β parametrivektori ja y havaintovektori. Residuaalivекtori e on $n + 1$ -vektori, jonka kukin alkio ilmaisee, kuinka paljon malli yliarvioi tai aliarvioi havainnon. Jos residuaalit ovat systemaattisesti positiivisia tai negatiivisia tietyillä selittävän muuttujan arvoilla, lineaarinen malli ei kuvaa dataa riittävän joustavasti. Tämä antaa perusteen paloittaisen regression käytölle. (Jekel & Venter, 2019, s. 1–5.)

Paloittaisessa lineaarisessa regressiossa selittävän ja selitettävän muuttujan suhde mallinnetaan useammalla erillisellä lineaarisella funktiolla eri arvoalueilla. Kohtaa, jossa funktio vaihtaa kulmakerrointa, kutsutaan murtopisteeksi. (Jekel ja Venter, 2019, s. 1.) Pokiya (2022, s. 4) määrittelee murtopisteet lineaaristen segmenttien katkaisupisteiksi, joissa yksi segmentti päättyy ja uusi alkaa. Yksinkertaisimmassa tapauksessa, jossa katkaisupiste sijaitsee kohdassa $x = c$, malli muodostuu kahdesta ehdollisesta lineaarifunktiosta

$$y = a_1 + b_1x, \quad x \leq c, \quad (66)$$

$$y = a_2 + b_2x, \quad x > c, \quad (67)$$

$$a_1 + b_1c = a_2 + b_2c. \quad (68)$$

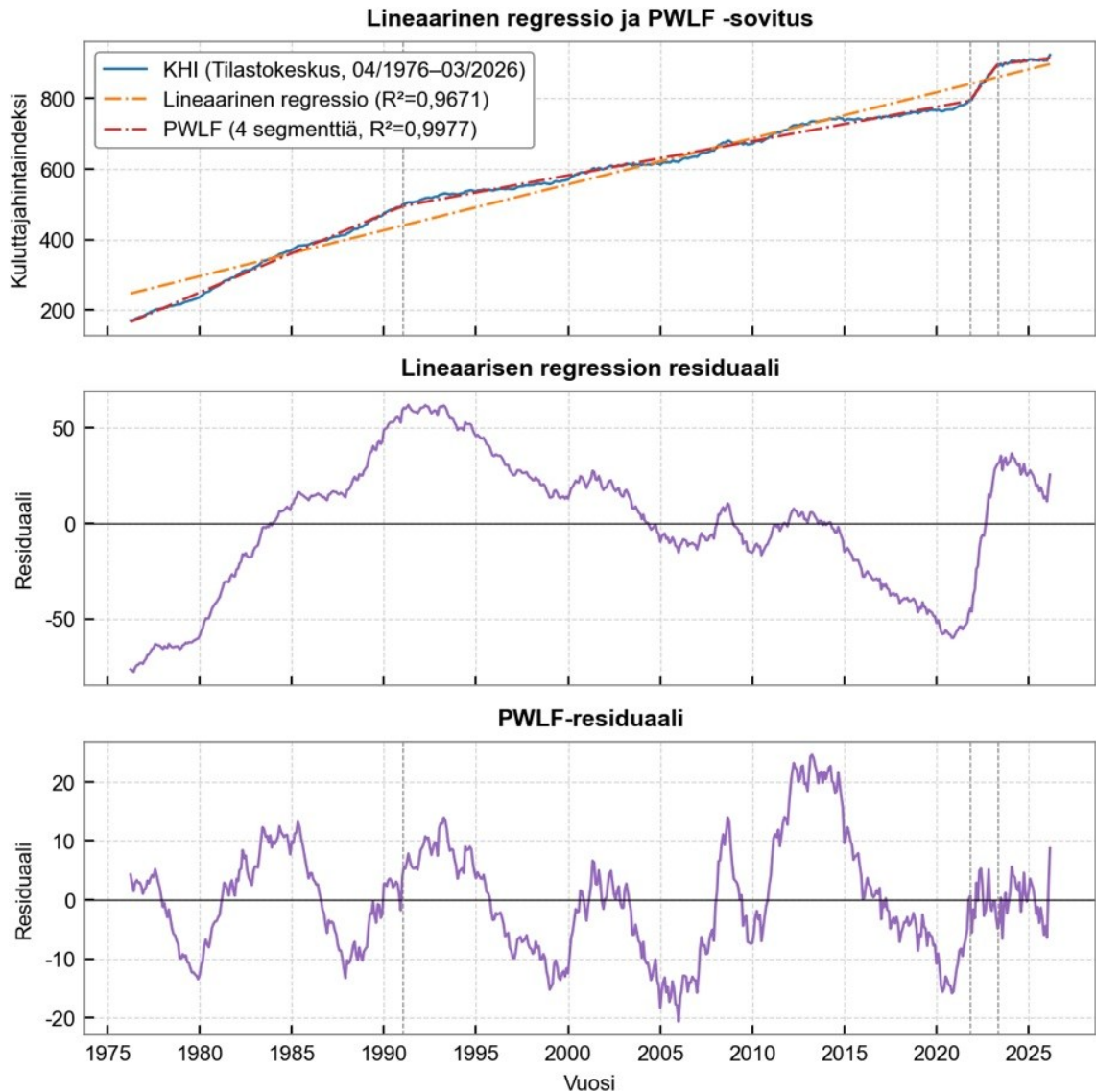
Kaavan (68) jatkuvuusehdon avulla malli voidaan esittää yhtenä yhtälönä dummy-muuttujan x_k avulla. Tällöin regressioyhtälö saa muodon

$$\hat{y} = b_0 + b_1x_1 + b_2(x_1 - x^{(k)})x_k, \quad x_k = \begin{cases} 0 & \text{jos } x_1 \leq x^{(k)} \\ 1 & \text{jos } x_1 > x^{(k)} \end{cases}, \quad (69)$$

missä x_1 on riippumattoman muuttujan arvo, $x^{(k)}$ on katkaisupisteen arvo ja x_k on katkaisupisteen dummy-muuttuja. (Pokia 2022, s. 6.) Kun murtopisteiden sijainnit ovat tuntemattomia, ne haetaan minimoimalla jäännösneliösumma. Jekel ja Venter (2019, s. 7) esittelevät tähän kaksi strategiaa: differentiaalievoluution sekä usean aloituspisteen gradienttipohjaisen LBFGS-algoritmin. Kumpikin muodostaa kaksoissisäkkäisen optimoinnin, jossa ulkokehä etsii optimaalisia katkaisupisteiden sijainteja ja sisäkehä ratkaisee pienimmän neliösumman β -parametrit. Menetelmä on toteutettu Python-ympäristöön kehitetyssä pwlf-kirjastossa, joka tuottaa myös β -parametrien standardivirheet ja p-arvon diagnostista arviointia varten (Jekel & Venter, 2019, s. 9).

Kuvio 15 esittää lineaarisen ja paloittaisen lineaarisen regression sovituksen KHI-aikasarjalle. Malli tunnistaa kolme murtopistettä, jotka jakavat aineiston neljään

segmenttiin selvästi poikkeavilla kulmakertoimilla. Selitysaste osoittaa, että pitkän aikavälin trendi kuvautuu erittäin tarkasti. Alempi osakuviot esittää trendistä erotetun residuaalin aikasarjan, joka toimii syötteenä CEEMDAN-algoritmile.



Kuvio 15. Kuluttajahintaindeksin lineaarinen ja paloittainen lineaarinen regressiosovitus.

5.3.2 CEEMDAN-algoritmi

Torres ym. (2011) esittävät joukkoistetun empiirisen CEEMDAN-moodihajotelmamenetelmän adaptiivisella kohinalla. CEEMDAN hajottaa signaalin $x[n]$ joukkoon oskillaatiokomponentteja, joita kutsutaan luontaisiksi moodifunktioiksi IMF.

Menetelmä perustuu siihen, että kullakin vaiheella muodostetaan yksi yhteinen residuaali, josta lasketaan seuraava IMF. (Torres ym., 2011, s. 4144.)

Olkoon $x[n]$ analysoitava signaali ja $w^i[n]$ kohinan realisaatio normaalijakaumasta $N(0, 1)$, missä $i = 1, \dots, I$ on kohinatoisto. Operaattori $E_j(\cdot)$ tuottaa annetun signaalin j :nnen EMD-moodin (engl. empirical mode decomposition). Ensimmäinen moodi saadaan keskiarvona, kun signaaliin lisätään pieni kohina ja kustakin tois-
tosta otetaan EMD-hajotelman ensimmäinen moodi (Torres ym., 2011, s. 4144.)

$$\widetilde{\text{IMF}}_1[n] = \frac{1}{I} \sum_{i=1}^I \text{IMF}_1^i[n], \quad (70)$$

jonka jälkeen lasketaan ensimmäinen yhteinen residuaali

$$r_1[n] = x[n] - \widetilde{\text{IMF}}_1[n]. \quad (71)$$

Seuraavat moodit määritetään iteratiivisesti lisäämällä kuhunkin residuaaliin adaptiivisesti skaalattu kohinakomponentti $\varepsilon_k E_k(w^i[n])$

$$\widetilde{\text{IMF}}_{k+1}[n] = \frac{1}{I} \sum_{i=1}^I E_1(r_k[n] + \varepsilon_k E_k(w^i[n])), \quad (72)$$

jonka jälkeen residuaalit päivitetään rekursiivisesti

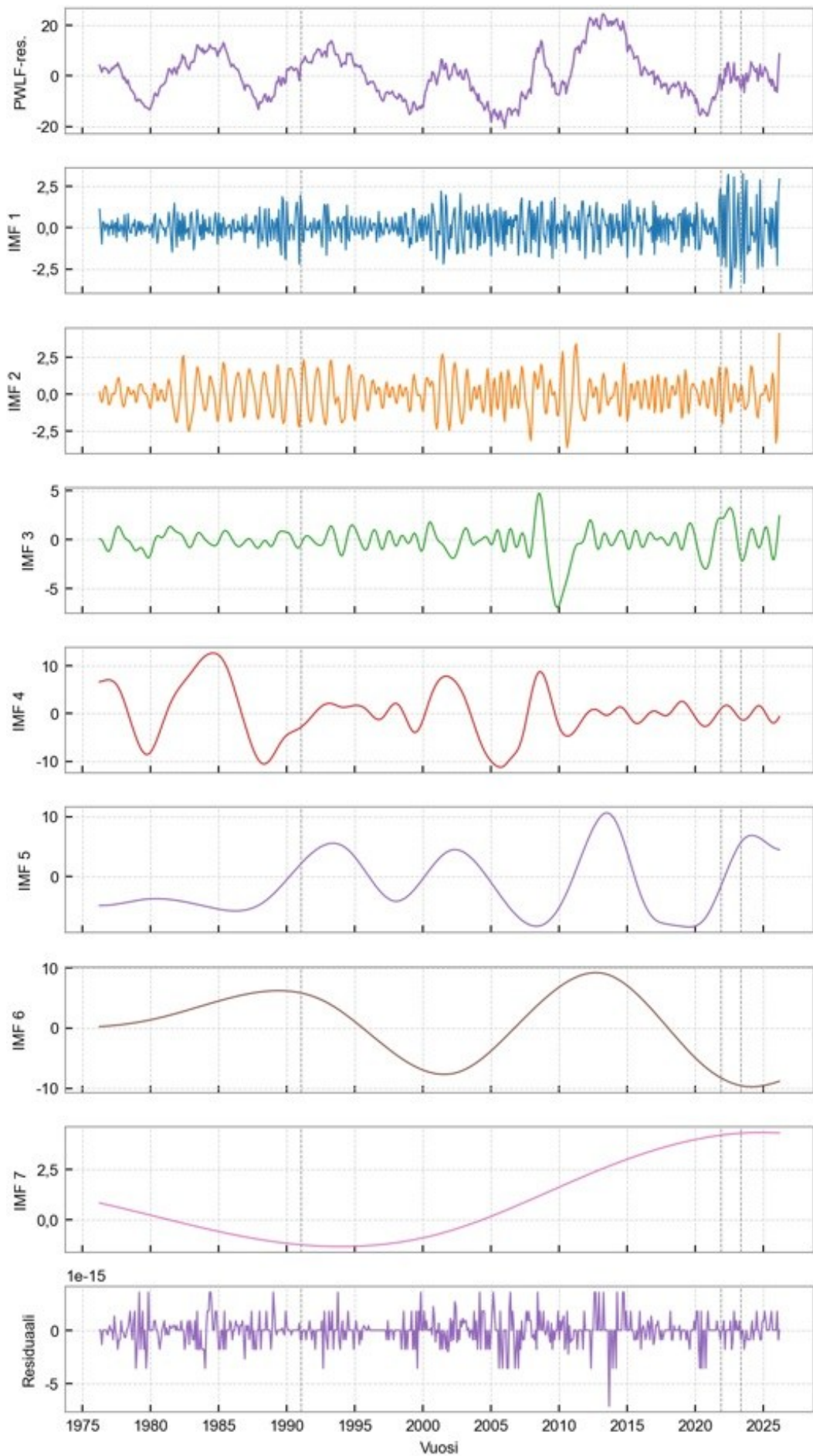
$$r_k[n] = r_{k-1}[n] - \widetilde{\text{IMF}}_k[n]. \quad (73)$$

Algoritmi pysähtyy, kun residuaalissa ei ole enää vähintään kahta ääriarvoa. Loppullinen rekonstruktio on täydellinen

$$x[n] = \sum_{k=1}^K \widetilde{\text{IMF}}_k - R[n], \quad (74)$$

missä $R[n]$ on residuaalijäännös. Kerroin ε_k mahdollistaa signaali-kohinasuhteen säätämisen kullakin hajotustasolla erikseen. Numeerinen rekonstruktiovirhe jää suuruusluokkaan 2×10^{-15} , mikä vastaa tietokoneen laskentatarkkuudesta johtuvaa pyöristysvirhettä. (Torres ym., 2011, s. 4144–4147.)

Kuvio 16 esittää PWLF-residuaalin CEEMDAN-hajotelman seitsemän IMF-komponenttia IMF 1–7 sekä loppujäännöksen. Komponentit järjestyvät taajuuden mukaan korkeimmasta matalimpaan: IMF 1–2 edustavat lyhyen aikavälin vaihtelua, IMF 3–4 keskipitkää syklistä ja IMF 5–7 pitkän aikavälin trendimäisiä oskillatioita. Loppujäännös on suuruusluokaltaan 10^{-15} , mikä osoittaa hajotelman olevan käytännössä täydellinen.



Kuvio 16. Kuluttajahintaindeksin PWLF-residuaalin CEEMDAN-hajotelma.

5.3.3 Lookback-ikkunan valinta

Kuluttajahintaindeksin pitkän aikavälin ennustaminen edellyttää, että mallin syötteelle asetetaan perusteltu kontekstiajan pituus eli lookback-ikkuna. Liian lyhyt ikkuna ei kykene hahmottamaan signaalin matalataajuisia syklejä, kun taas liian pitkä ikkuna kasvattaa epäolennaisen historian painoarvoa, lisää parametrimäärää ja hidastaa laskentaa. Inflaatiomallissa lookback-ikkuna arvioidaan automaattisesti Hilbert-muunnoksella CEEMDAN-hajotuksen tuottamista sisäisistä moodikomponenteista.

Hilbert-muunnosta käytetään sen hetkellisen taajuussisällön määrittämiseen. Jokaisista IMF:ää käsitellään ikään kuin se olisi todellinen värähtely, joka muutetaan analyttiseksi signaaliksi eli kompleksiseksi signaaliksi, jonka reaaliosa on IMF ja jonka imaginaariosa on IMF:n Hilbert-muunnos. Hilbert-muunnos $Hc_k(t)$ on signaalin 90-asteen vaiheensiirto, mikä mahdollistaa $c_k(t)$:n esittämisen muodossa

$$z_k(t) = c_k(t) + i H\{c_k(t)\} = A_k(t) e^{i\theta_k(t)}, \quad (75)$$

missä $A_k(t)$ on hetkellinen amplitudi ja θ_k on k :nnen hetkellinen vaihekulma (Huang ym. 1998, s. 9–10). Hetkellinen kulmataajuus määritellään vaihekulman aikaderivaattana

$$\omega_k(t) = \frac{d\theta_k(t)}{dt}. \quad (76)$$

IMF-luokka on määritelty siten, että hetkellinen taajuus on mielekäs kaikilla ajanhetkillä: nollakohtien ja ääriarvojen lukumäärät ovat yhtä suuret tai eroavat enintään yhdellä, ja signaali on paikallisesti symmetrinen ylä- ja alavaippakäyriensä suhteen (Huang ym. 1998, s. 13). Liitteessä 2 esitetään PWLF-residuaalin CEEMDAN-hajotelman seitsemän IMF-komponenttia hetkellisine jaksoineen ja amplitudiverhokäyrineen. Hajotelman dominoivan syklin muodostavat IMF 4 ja IMF 5, joiden hetkelliset jaksot osuvat referenssiperiodin tasolle.

Dominoivat sykliset komponentit tunnistetaan Hilbertin marginaalispektrin avulla, joka saadaan integroimalla Hilbertin amplitudispektri $H(\omega, t)$ ajan yli (Huang ym. 1998, s. 26–27). Spektri $H(\omega, t)$ kuvaa IMF-komponenttien hetkellisten amplitudien jakaumaa aika-taajuustasossa, ja sen aikaintegraali muodostaa marginaalispektrin

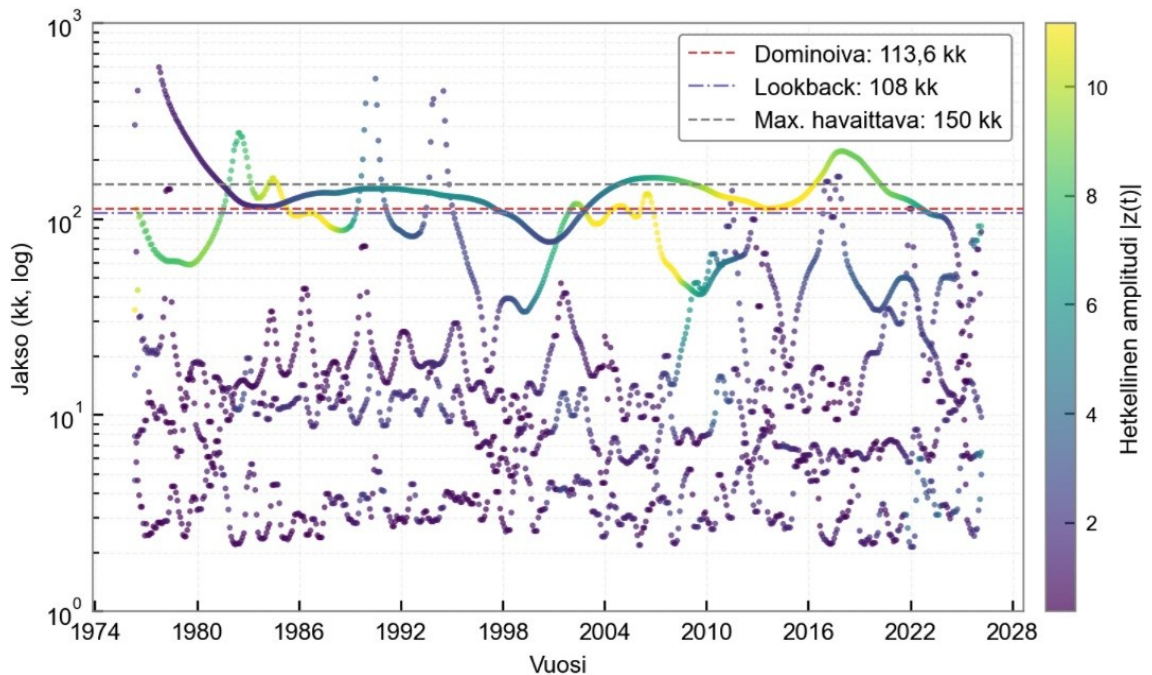
$$h_i(\omega) = \int_0^T H(\omega, t) dt. \quad (77)$$

Marginaalispektri $h_i(\omega, t)$ kuvaa kunkin taajuuden kumulatiivista amplitudia havaintojakson yli (Huang ym. 1998, s. 27–28). Korkea $h_i(\omega, t)$ kertoo, että kyseisen taajuuden värähtely on esiintynyt aineistossa joko pitkäkestoisesti tai voimakkaalla amplitudilla. Inflaatiomallissa dominoiva sykli määritellään marginaalispektrin globaalina maksimina ω^* , ja sen käänteisluku P_{ref} toimii lookback-ikkunan referenssiiperiodina

$$\omega^* = \arg \max_{\omega} h_i(\omega), \quad (78)$$

$$P_{\text{ref}} = \frac{2\pi}{\omega^*}. \quad (79)$$

Kuvio 17 esittää valittujen moodifunktioiden Hilbert-spektrin aika-jaksotasossa. Jokainen piste vastaa yhden IMF:n hetkellistä jaksoa. Värin kirkkaus ilmaisee hetkellisen amplitudin $|z(t)|$ voimakkuuden. Katkoviivat osoittavat kolme tulkintarajaa. Hallitseva jakso 113,6 kk on aikasarjan energiapainotettu päävärähtely, tarkasteluikkuna 108 kk on analyysin konteksti-ikkuna ja suurin havaittava jakso 150 kk on pisin luotettavasti erotettavissa oleva jakso.



Kuvio 17. Valittujen IMF:ien Hilbert-spektri aika-jaksotasossa.

5.3.4 Mallin koulutus ja validointi

Walk-forward-validoinnissa koulutusikkunaa laajennetaan tai liu'utetaan yksi tai useampi jakso kerrallaan niin, että malli koulutetaan aina vain historiallisella datalla suhteessa validointipisteeseen (Lewinson, 2022, s. 206). Aikasarjaennustamisen suorituskyvyn arviointimenetelmät jaetaan kahteen päätyyppiin, out-of-sample-menetelmiin ja ristiinvalidointiin, jotka säilyttävät havaintojen ajallisen järjestyksen. Yksittäiseen jakokohtaan perustuva arviointi on validoinnin keskeinen haaste, sillä yhdestä origosta saatu virhe-estimaatti on herkkä jakokohdan ominaispiirteille ja voi vääristää käsitystä mallin yleistyskyvystä. Luotettavamman estimaatin saamiseksi arviointi suositellaan toistettavaksi useassa eri testausikkunassa. (Cerqueira ym., 2020, s. 1999–2003, 2019.)

Walk-forward-validoinnissa ajallinen järjestys säilyy aina koulutus- ja testijoukon välillä, eikä mallia testata koulutusjoukkoa edeltäneillä havainnoilla (Cerqueira ym., 2020, s. 1999). Menetelmästä erotetaan kaksi päämuotoa: laajenevassa ikkunassa (engl. expanding tai growing window) aiemmat lohkot liitetään koulutusjoukkoon kierros kierrokselta, kun taas liukuvassa ikkunassa (engl. sliding window) vanhimmat lohkot pudotetaan pois uusien tieltä, mitä suositetaan erityisesti epästationaarisissa ympäristöissä, joissa vanha data vanhenee suhteessa nykytilaan (Cerqueira ym., 2020, s. 2001).

Laajenevan ikkunan Walk-forward-menetelmässä koulutusjoukko kasvaa kierros kierrokselta, ja validointi-ikkuna liikuu askelkoolla kohti aikasarjan loppua. Olkoon Y pituudeltaan T oleva aikasarja havainnoista y_1, y_2, \dots, y_T , joille $y_i \in \mathbb{R}$ kaikilla $i \in \{1, \dots, T\}$. Aineisto jaetaan kahtia niin, että aikasarjan loppuun varataan pituudeltaan T_h oleva pidätetty out-of-sample-testijakso, jota ei käytetä mallin kehittämiseen eikä hyperparametrien valintaan. Jäljelle jäävä walk-forward-validoinnin aineisto on pituudeltaan

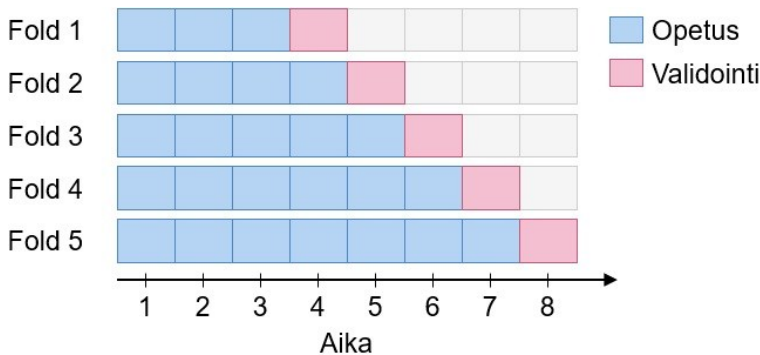
$$T_{wf} = T - T_h. \quad (80)$$

Validointimenettelyn parametrit ovat alkuperäisen koulutusikkunan koko n , askelkoko s , validointihorisontti h sekä datanpoisto p koulutus- ja validointijakson välillä. Foldien eli validointikierrosten kokonaismäärä K saadaan ehdolla

$$K = \left\lfloor \frac{T_{wf} - n - p - h}{s} \right\rfloor + 1. \quad (81)$$

Kaava (81) ilmaisee, kuinka monta kertaa ikkuna ehtii edetä askelkoolla s ennen kuin jäljelle jäävä aineisto ei enää riitä datanpoistolle ja sitä seuraavalle horisontin mittaiselle validointijaksolle. Datanpoisto vaimentaa aikasarjan autokorrelaatiosta syntyvää tietovuotoa koulutus- ja validointijakson rajalla, mikä vähentää validointivirheen harhaa yleistyskyvyn mittarina.

Laajenevan ikkunan rakennetta havainnollistaa kuvio 18, joka esittää viiden kierroksen validointimenettelyn kahdeksan aikajakson sarjalle. Koulutusjoukko kasvaa kierros kierrokselta ja validointipiste siirtyy seuraavaan ajanjaksoon, jolloin myöhemmillä kierroksilla malli hyödyntää laajempaa historiallista aineistoa kuin alkuvaiheessa. Lewinsonin (2022, s. 207) mukaan tämä on menetelmän keskeinen rajoitus, koska aikaisempien kierrosten virhemitat eivät ole vertailukelpoisia myöhempien kanssa, koska niukalla aineistolla malli ei ehdi oppia kausivaihtelua.



Kuvio 18. Walk-forward-validointi laajenevalla ikkunalla (Lewinson, 2022, s. 207, muokattu).

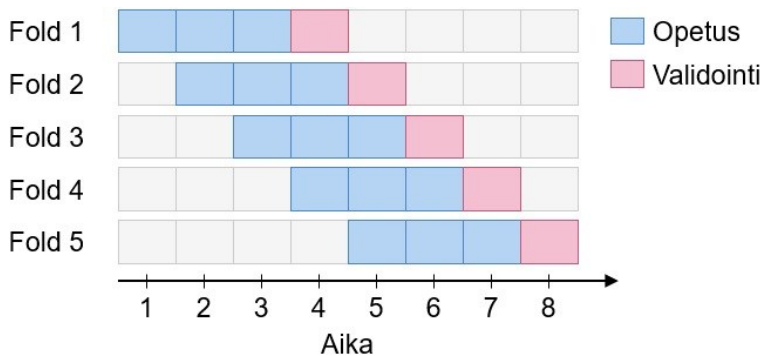
Liukuvan ikkunan menetelmässä koulutusikkunan koko pysyy kiinteänä parametrimina w , jolloin ikkunaa siirretään eteenpäin jokaisen kierroksen jälkeen poistamalla vanhin havainto s ja lisäämällä uusin. Kierroksella k koulutusjoukko ja validointijoukko määritellään

$$S_k^{\text{koulutus}} = \{y_{s(k-1)+1}, y_{s(k-1)+2}, \dots, y_{s(k-1)+w}\}, \quad (82)$$

$$S_k^{\text{validointi}} = \{y_{s(k-1)+w+p+1}, y_{s(k-1)+w+p+2}, \dots, y_{s(k-1)+w+h}\}. \quad (83)$$

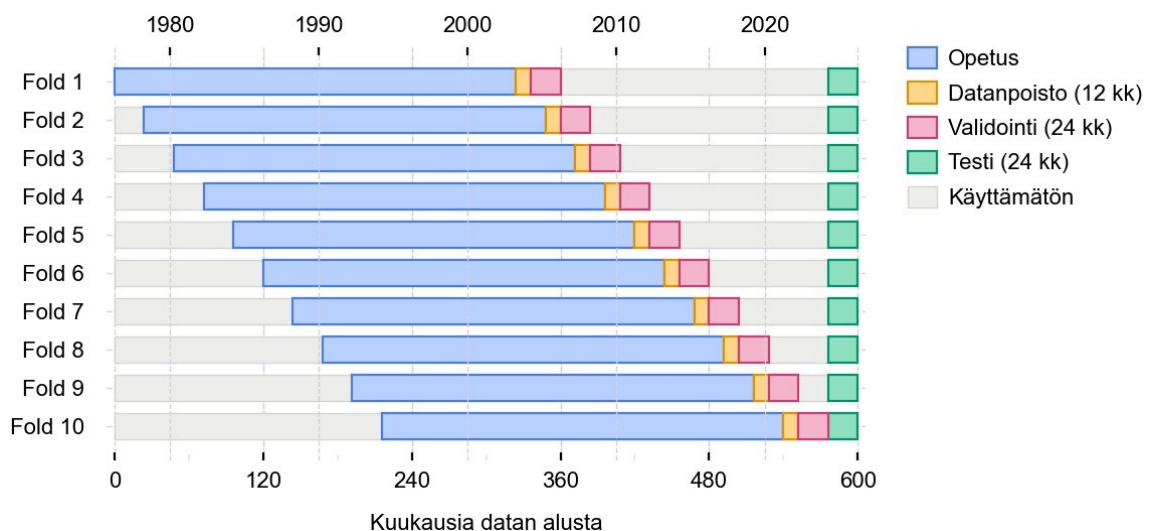
Liukuvan ikkunan rakennetta havainnollistaa kuvio 19, joka esittää viiden kierroksen validointimenettelyn kahdeksan aikajakson sarjalle. Ensimmäinen kierros on

identtinen laajenevan ikkunan kanssa. Mutta sitä seuraavilla kierroksilla koulutusaineiston koko pysyy vakiona, jolloin kierrosten virhemitat ovat suoraan vertailukelpoisia. Lewinsonin (2022, s. 207) mukaan menetelmä soveltuu erityisesti tilanteisiin, joissa koulutusaineistoa on runsaasti tai joissa kaukainen historia ei ole olennainen tulevien arvojen ennustamisen kannalta.



Kuvio 19. Walk-forward-validointi liukuvalla ikkunalla (Lewinson, 2022, s. 208, muokattu).

Kuvio 20 esittää inflaatiomallin liukuvan ikkunan walk-forward-validoinnin rakenteen. Aineisto jaetaan ajallisesti peräkkäisiin osiin, jolloin malli ei näe tulevaisuuden havaintoja koulutusvaiheessa. Kullakin kierroksella 324 kuukauden koulutusikkuna liikuu 24 kuukautta eteenpäin, jolloin vanhimmat havainnot korvautuvat uusilla. Koulutuksen ja validoinnin välissä on 12 kuukauden datanpoisto, joka estää tietovuodon, ja sitä seuraava 24 kuukauden validointijakso vastaa mallin ennustehorisonttia. Viimeiset 24 kuukautta aikasarjasta on varattu testijaksoksi lopullisen mallin arviointiin.



Kuvio 20. Inflaatiomallin Walk-Forward-validoinnin periaate liukuvalla ikkunalla.

Jokaisen foldin validointitulokset lasketaan kolmella toisiaan täydentävällä virhemetriikalla. MAE mittaa keskimääräistä absoluuttista virhettä alkuperäisellä asteikolla. RMSE painottaa neliöinnin kautta suuria poikkeamia voimakkaammin, ja MASE suhteuttaa mallin virheen naiivin satunnaiskulkuennusteen virheeseen, jolloin yhtä pienempi arvo osoittaa parannuksen vertailukohtaan nähden. Metriikat lasketaan kaavoilla (84), (85) ja (86) seuraavasti

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (84)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (85)$$

$$\text{MASE} = \frac{\text{MAE}}{\text{MAE}_{\text{naiivi}}}, \quad \text{MAE}_{\text{naiivi}} = \frac{1}{n-1} \sum_{t=2}^n |y_t - y_{t-1}|. \quad (86)$$

Naiivin vertailuennusteen virhe $\text{MAE}_{\text{naiivi}}$ lasketaan kertaalleen koko walk-forward-aineistolle, ja samaa nimittäjää käytetään kaikissa foldeissa. Tämä yhteinäistää MASE-arvojen tulkinnan kierrosten välillä, sillä vaikka aikasarjan absoluuttinen taso tai vaihteluväli muuttuisivat pitkällä aikavälillä, MASE-arvot pysyvät keskenään vertailukelpoisina ja ilmaisevat johdonmukaisesti mallin suhteellisen paremmuuden vertailukohtaan nähden.

Walk-forward-validointi tuottaa lisäksi ennustehorisonttikohtaisen epävarmuusarvion. Kullekin ennusteaskeleelle kerätään virhejämmät kaikista foldeista, ja niiden perusteella lasketaan askelkohtainen empiirinen keskihajonta. Tämä hajonta kuvaa mallin tyypillistä virhettä kyseisellä ennustepituudella. Näin saatu mallivirhekomponentti σ_{malli} yhdistetään kahteen analyttisesti johdettuun komponenttiin, trendin ja historian epävarmuuteen, jolloin muodostuu kokonaisepävarmuusarvio

$$\sigma_{\text{kok}}(t) = \sqrt{\sigma_{\text{malli}}^2(t) + \sigma_{\text{trendi}}^2(t) + \sigma_{\text{historia}}^2(t)}, \quad (87)$$

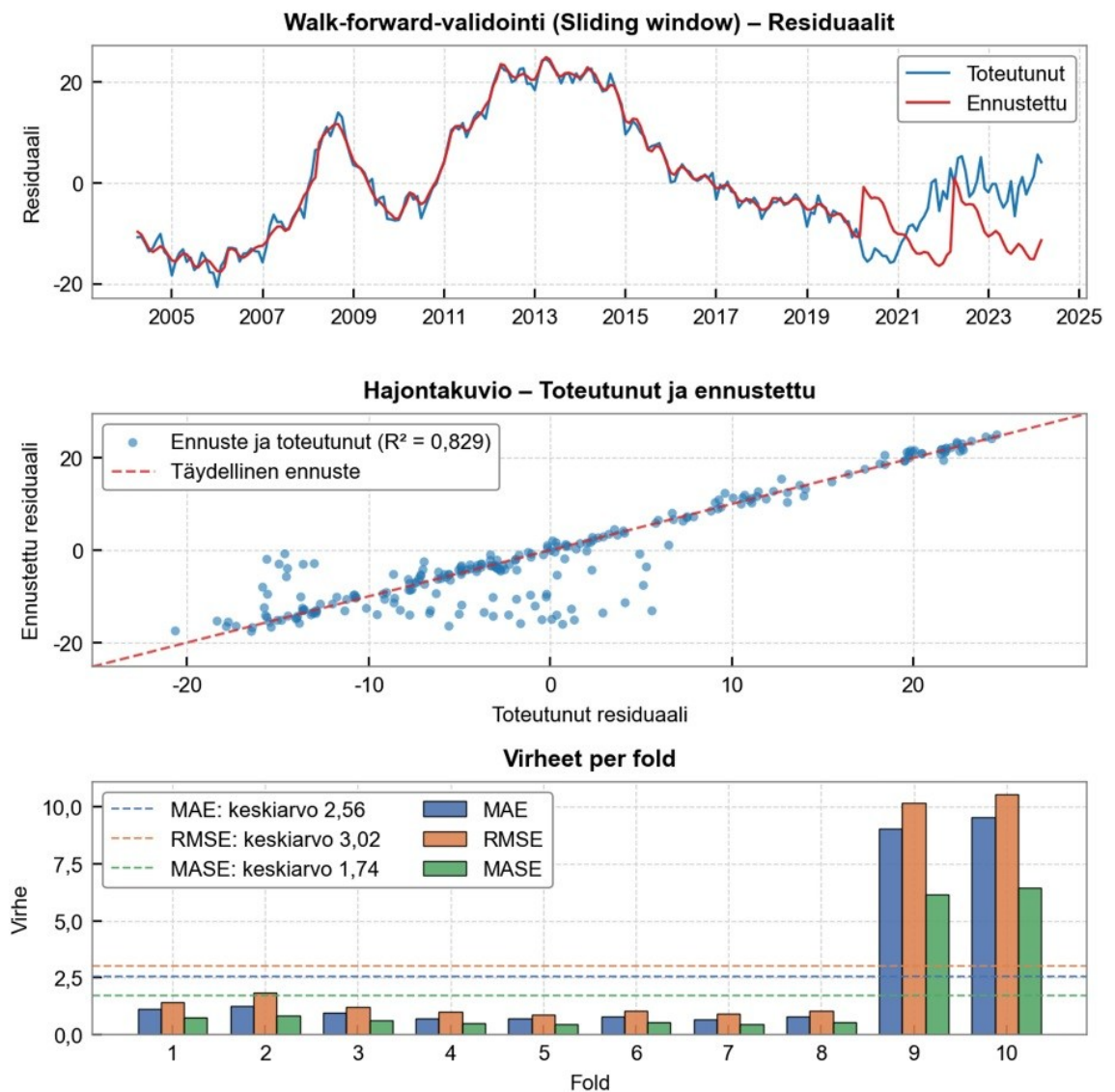
jossa $\sigma_{\text{kok}}(t)$ on ennusteen kokonaisepävarmuus t kuukautta ennustejakson alusta. Trendin epävarmuus kasvaa ennustehorisontin pidentyessä lineaarisesti pitkän aikavälin trendisovituksen kulmakertoimen keskivirheen σ_{β} skaalaamana

$$\sigma_{\text{trendi}}(t) = \sigma_{\beta} \times t. \quad (88)$$

Historiallisten residuaalien ennakoimaton osa kasautuu satunnaiskulkuoletuksen mukaisesti ajan neliöjuureen verrannollisesti. Koska $\sigma_{\text{residuaali}}$ on määritetty vuositasolle, aika skaalataan vuosiksi jakamalla se kahdellatoista kaavalla

$$\sigma_{\text{historia}}(t) = \sigma_{\text{residuaali}} \times \sqrt{\frac{t}{12}}. \quad (89)$$

Kuvio 21 esittää validoinnin tulokset. Malli seuraa residuaalien yleistä kehityssuuntaa keskimääräisellä MAE-arvolla 2,56 ja RMSE-arvolla 3,02 KHI-pistettä. Suurimmat virheet havaitaan foldeissa 9 ja 10, jotka ajoittuvat koronapandemian vuosiin 2020–2023 ja poikkeavat rakenteellisesti aikasarjan aiemmasta dynamiikasta. Hajontakuvion korkea selitysaste osoittaa mallin kattavan lähes koko residuaalivaihtelun 24 kuukauden ennustehorisontilla.

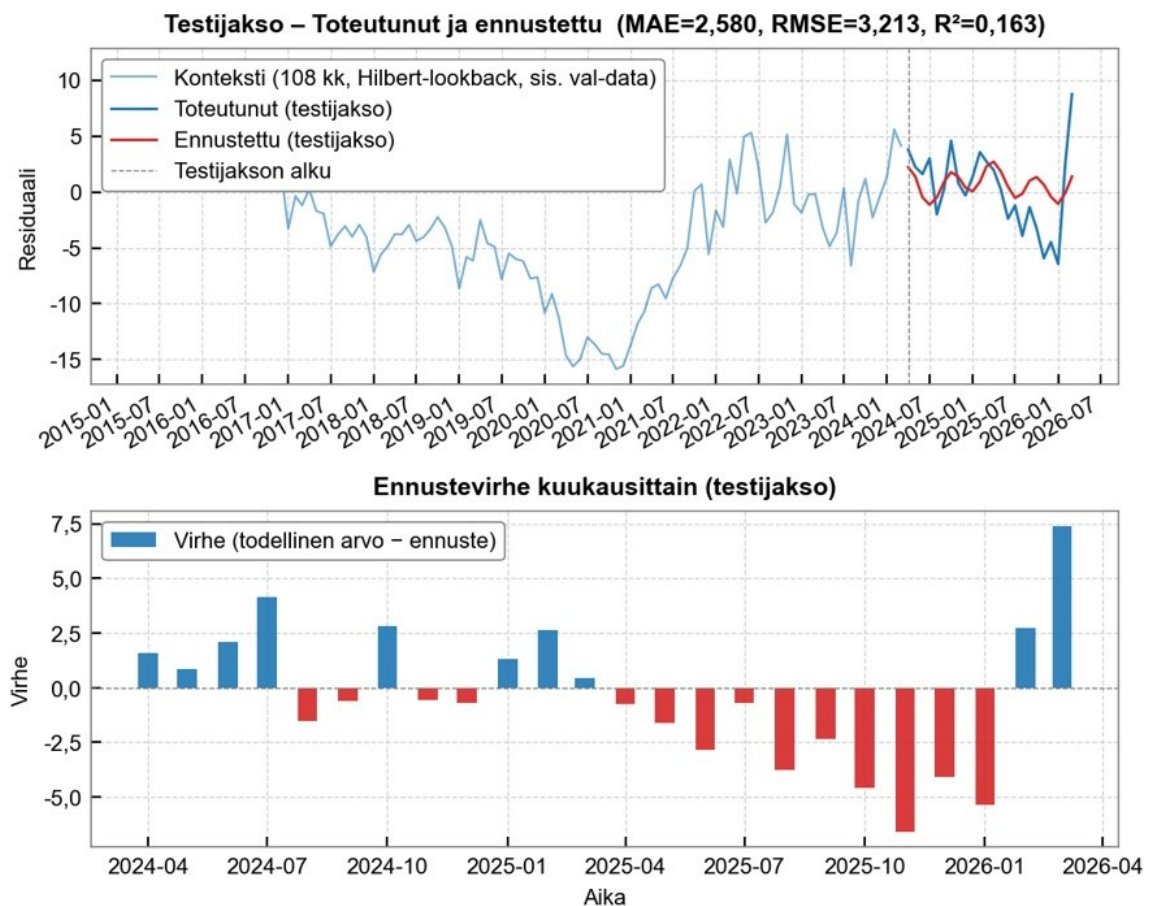


Kuvio 21. Liukuvan ikkunan Walk-Forward-validoinnin tulokset.

Liitteessä 3 esitetään gradienttien käyttäytyminen ja koulutusmetriikat koulutuksen aikana. Suurimmat gradientit keskittyvät $\theta_{f_fc.weight}$ -kerroksiin, joiden kautta lohkojen ennusteet välittyvät verkossa eteenpäin, eikä häviävän pieniä gradientteja havaita. Alun piikki tasoittuu nopeasti, ja arvot vakiintuvat maltilliselle tasolle ilman räjähtämistä tai häviämistä. Validointijoukon virhemittarit laskevat tasaisesti. Myös backcast- ja forecast-häviön käyrät seuraavat toisiaan ilman erkanemista. Tulokset osoittavat koulutuksen etenevän vakaasti ja mallin oppivan yleistettäviä ennustepiirteitä ilman merkittävää ylisovittamista.

5.3.5 Ennusteet

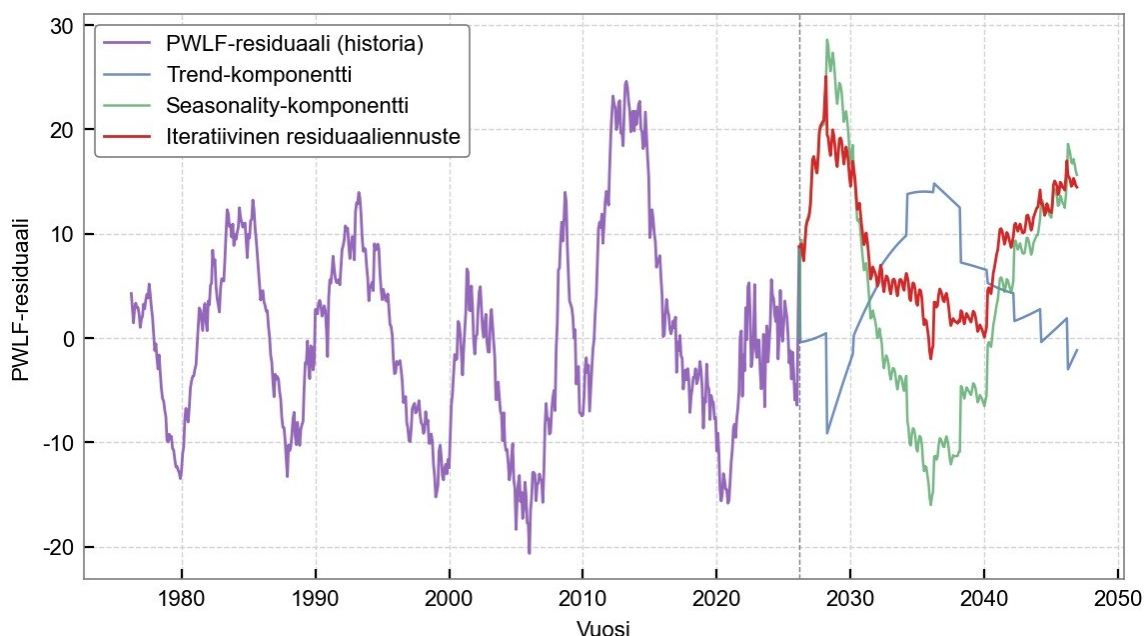
Validoinnin jälkeen mallin lopullinen suorituskky arvioidaan testijaksolla, johon kuuluvat aikasarjan viimeiset 24 kuukautta eli huhtikuu 2024 – maaliskuu 2026. Tämä jakso pidetään täysin erillään koulutuksesta ja walk-forward-validoinnista, ja vasta valmis malli ennustaa sen kerralla. Tulos havainnollistaa, miten malli suoriutuu aidosti tuntemattomalla datalla (kuvio 22).



Kuvio 22. KHI-pisteluvun walk-forward-ennuste 24 kuukauden testijaksolle.

Testikauden ennuste arvioi koulutetun mallin suorituskykyä täysin koskemattomalla testijaksolla, jota malli ei ole nähnyt koulutuksen eikä validoinnin aikana. N-BEATS tuottaa tälle jaksolle residuaaliennusteen, joka yhdistetään PWLF- tai lineaariseen trendisovitukseen lopulliseksi KHI-ennusteeksi. Ennustetta verrataan toteutuneisiin arvoihin (liite 4). Liitteen 4 kuviossa esitetään myös N-BEATS-mallin tuottamat trendi- ja kausittaisuuskomponentit, jotka näyttävät, miten ennuste jakautuu hitaaseen trendisiirtymään ja nopeampaan kausivaihteluun.

Kuluttajahintaindeksin pitkän aikavälin ennuste muodostetaan komposiittimallina, jossa aikasarjan dynamiikka jaetaan kahteen erikseen estimoituun komponenttiin, paloittain lineaariseen deterministiseen trendiin ja N-BEATS-neuroverkolla tuotettuun stokastiseen residuaaliin. Tämä rakenne erottaa indeksin pitkän aikavälin tason syklisestä vaihtelusta, jolloin kumpaakin komponenttia voidaan mallintaa sen luonteeseen parhaiten soveltuvalla menetelmällä. Residuaalikomponentti on se trenditön jäännös, jonka PWLF jätti jäljelle ja jota CEEMDAN- ja Hilbert-analyysi ovat luonnehtineet. Tälle komponentille N-BEATS tuottaa itsenäisen ennusteen (kuvio 23).



Kuvio 23. N-BEATS-mallin tuottama PWLF-residuaalin pitkän aikavälin ennuste.

Pitkän aikavälin residuaaliennuste tuotetaan iteratiivisella moniaskeleisella ennustamisella, jossa N-BEATS-mallia sovelletaan toistuvasti koulutushorisontin mittaisiin askelin. Malli on koulutettu 24 kuukauden ennustehorisonttiin, sillä se on

rakennettu tuottamaan kerralla enintään tämän mittainen ennuste, eikä suora 240 kuukauden ekstrapolointi siksi ole mahdollista. Kukin iteraatio ennustaa seuraavat 24 kuukautta Hilbert-analyysin määrittämästä konteksti-ikkunasta, minkä jälkeen ennustetut arvot liitetään historiaan ja seuraava iteraatio alkaa päivitetystä ikkunasta. Koko 20 vuoden horisontti katetaan näin kymmenellä peräkkäisellä iteraatiolla.

Lopullisen ennusteen luottamusvälit muodostetaan sadan toiston Monte Carlo -simulaatiolla, jossa koko ennusteputki ajetaan satunnaiskohinalla. Kukin simulaatio vaihtelee neljää epävarmuuslähdeä toisistaan riippumattomasti. Trendin kulmakerrointa vaihdellaan sen keskivirheen mukaisella normaalijakautuneella poikkeamalla, N-BEATS-mallin parametreja Monte Carlo Dropoutilla ja historiallista volatilitteettia kumulatiivisena satunnaiskulkuna. Neljäs lähde on walk-forward-validoinnista bootstrap-näytteistetty virhetermi. Empiiriset 50, 80 ja 95 prosentin luottamusvyöhykkeet lasketaan simulaatiojakauman persentileistä.

N-BEATS-mallin parametripävarmuus sisällytetään simulaatioon Monte Carlo Dropout (MCDO) -tekniikalla. MCDO perustuu neuroverkkojen dropout-menetelmään, jota käytetään ylisovittamisen ehkäisyyn koulutusvaiheessa. Tavanomaisessa käytössä dropout kytketään pois ennustevaiheessa, mutta MCDO:ssa se pidetään aktiivisena myös ennusteita laskettaessa. Gal ja Ghahramani (2016) osoittivat, että tämä menettely vastaa likimäärin bayesiläistä päättelyä. Jokainen eteenpäinajo tuottaa hieman erilaisen ennusteen, koska eri neuronit kytkeytyvät satunnaisesti pois käytöstä, ja näin muodostuva ennustejakauma heijastaa mallin parametripävarmuutta. Menetelmän etuna on, että se tuottaa epävarmuusarvioita ilman painojakaumien eksplisiittistä estimointia. (Gal & Ghahramani, 2016, viitattu lähteessä Kummaraka & Srisuradetchai, 2025, s. 3.)

Käytännössä jokainen Monte Carlo -kierros suorittaa N-BEATS-mallin täyden iteraatiivisen ennusteen dropoutin ollessa aktiivisena, jolloin satunnaisesti poiskytketyvät neuronit tuottavat kullakin kierroksella toisistaan poikkeavan kulkureitin. Simulaatioiden tuottamien ennusteiden $\{y_1^*, y_2^*, \dots, y_M^*\}$ empiirinen keskiarvo

$$\bar{y}^* = \frac{1}{M} \sum_{m=1}^M y_m^* \quad (90)$$

toimii pistemäisenä ennusteena. Ennusteiden varianssi

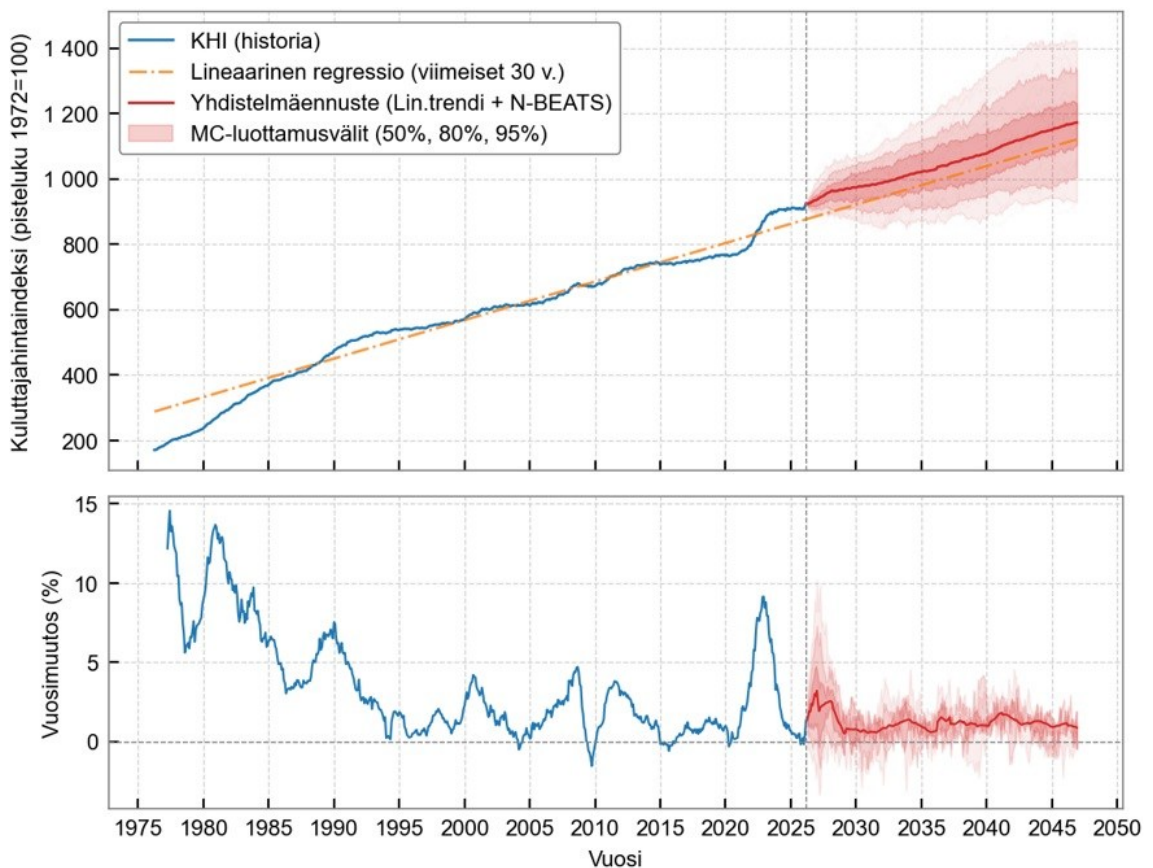
$$\sigma_{\bar{y}^*}^2 = \frac{1}{M} \sum_{m=1}^M (y_m^* - \bar{y}^*)^2 \quad (91)$$

kuvaa mallin ennuste-epävarmuutta. Näiden perusteella voidaan muodostaa 95 %:n luottamusvälin esitysmuoto

$$\bar{y}^* \pm 1,96 \sqrt{\sigma_{\bar{y}^*}^2}, \quad (92)$$

joka antaa arvion ennusteen todennäköisestä vaihteluvälistä. (Kummaraka & Srisuradetchai, 2025, s. 4.)

Kuviossa 24 ennuste osoittaa kuluttajahintaindeksin nousevan maltillisesti pitkällä aikavälillä. Piste-ennuste arvioi vuosittaiseksi inflaatioksi noin 1–2 % vuoteen 2046 asti, mikä on linjassa Euroopan keskuspankin keskipitkän aikavälin inflaatiotavoitteen 2 % kanssa. Luottamusvyöhykkeet 50 %, 80 % ja 95 % laajenevat odotetusti ajan myötä heijastellen kumuloituvaa epävarmuutta.



Kuvio 24. Kuluttajahintaindeksin pitkän aikavälin ennuste vuoteen 2046 asti.

Taulukko 2 kokoaa yhteen mallin keskeisten vaiheiden menetelmät, niiden tehtävät sekä keskeisimmät parametrit.

Taulukko 2. Inflaatiomallin menetelmäarkkitehtuurin vaiheet ja parametrit.

Vaihe	Menetelmä	Tulkinta
1. Trendin estimointi	Paloittainen lin. regressio (PWLF)	4 segmenttiä, SSR-minimointi
2. Signaalihajotelma	CEEMDAN-algoritmi	trials = 200, $\varepsilon = 0,05$, 7 IMF-komponenttia ja jäännös
3. Konteksti-ikkunan optimointi	Hilbert-muunnos (marginaalispektri)	Vähintään 2,5 havaittavaa sykliä
4. Ennustaminen	N-BEATS-neuroverkko (tulkittava)	2 pinoa (trendi + kausittaisuus) \times 3 lohkoa \times 4 kerrosta, leveydet [128, 512], ennustehorisontti 24 kk, dropout = 0,15
5. Validointi	Walk-Forward (liukuva ikkuna ja testi-kausi)	Koulutusikkuna 324 kk (27 v.), horisontti 24 kk, askel 24 kk, puhdistusväli 12 kk, testi 24 kk
6. Epävarmuus	Monte Carlo -simulaatio	100 simulaatiota, 4 häiriölähdettä (trendin kulmakerroin, MC Dropout, historiallinen volatilitteetti, WF-virhe bootstrap)

5.4 Pitkän aikavälin investointisuunnitelma

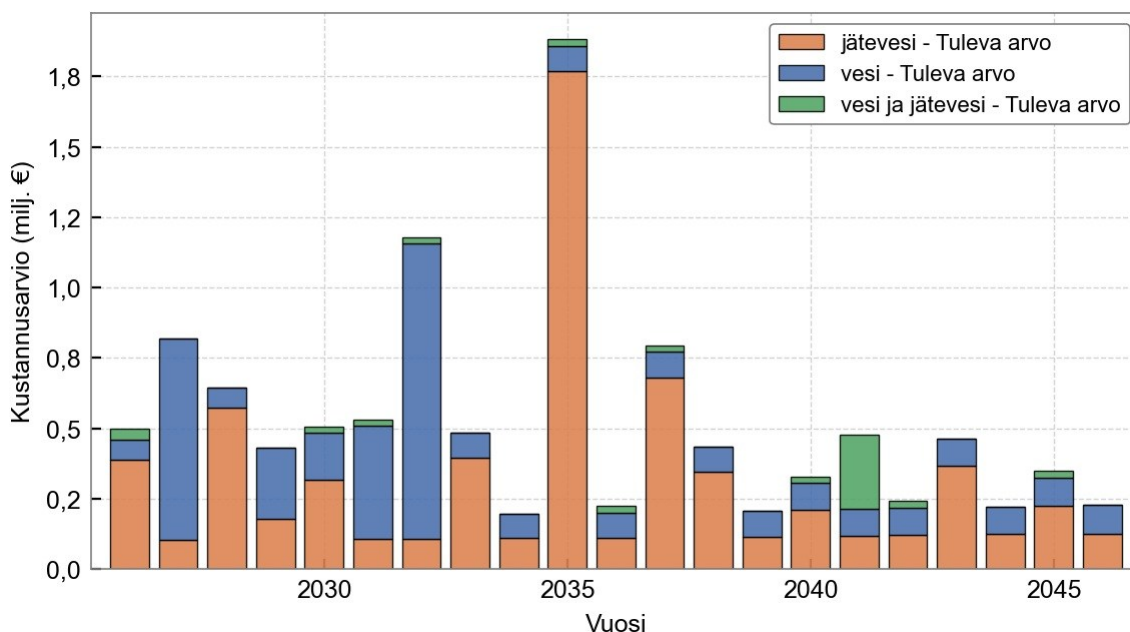
Pitkän aikavälin taloussuunnittelun keskeisiä haasteita ovat inflaation vaikutus suunniteltujen investointien kustannuksiin, käyttöomaisuuden taseen hallinta poistojen kautta sekä ulkoisen rahoituksen optimointi. Omaisuudenhallintamalli automatisoi laskentaprosessit ja tuottaa päätöksentekoa tukevia visualisointeja. Ohjelma käsittelee 20 vuoden mittaista investointisuunnittelua kuluva vuodesta alkaen, ja sen lähtötietoina toimivat investointitaulukko sekä N-BEATS-mallilla ennustetut kuluttajahintaindeksin kuvaamat inflaation vuosimuutokset.

Investointien kustannukset on arvioitu kuluva vuoden rahanarvossa eli nykyarvossa. Jotta tulevien investointivuosien kustannukset voidaan esittää toteutumisaikankohdan hintatasossa, ne muunnetaan tulevaksi arvoksi diskontauskaavalla

$$FV = PV \times \prod_{t=1}^T (1 + r_t), \quad (93)$$

missä FV on tuleva arvo, PV on nykyarvo, r_t on vuoden t inflaatioaste ja T on tarkasteluvuosien lukumäärä nykyhetkestä investointivuoteen. Kumulatiivinen

diskontauskorotus lasketaan nykyisestä vuodesta investointivuoteen saakka. Investointien esittely tiivistetään visualisoinniksi, joka kuvaa tulevien investointien jakautumista kustannuspaikoittain vesi-, jätevesi- ja yhteishankkeisiin vuosittain (kuvio 25).

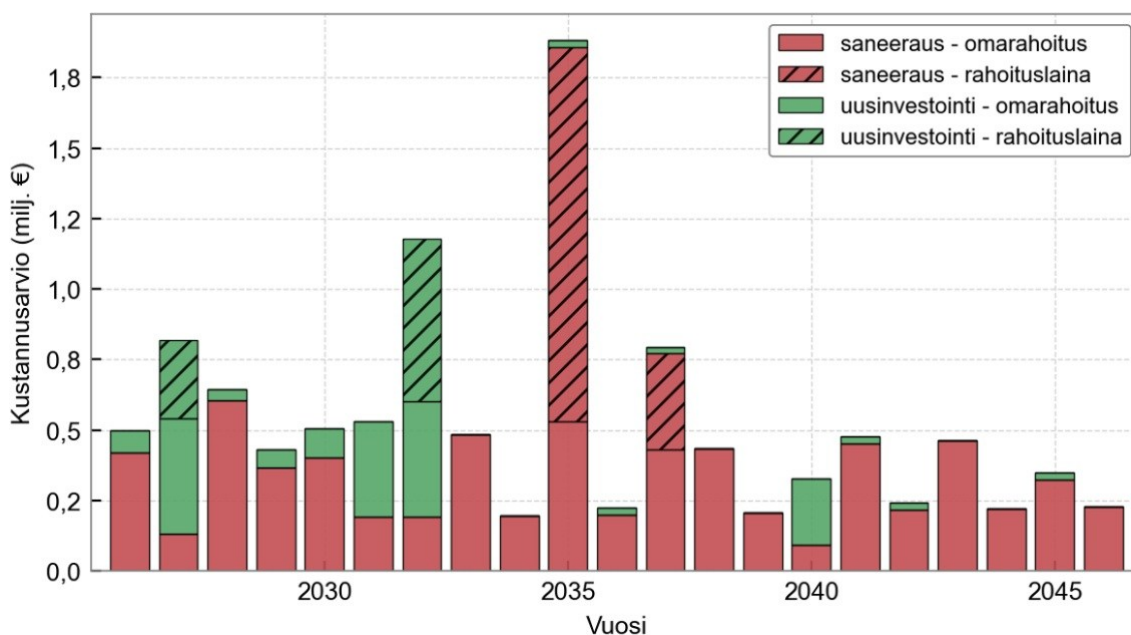


Kuvio 25. Case-vesihuoltolaitoksen investointikustannukset kustannuspaikoittain erään investointiskenaarion mukaisesti.

Pitkävaikutteiset investoinnit ja velkaantumispäätökset tuovat yrityksen tuloslaskelmaan lisää kiinteitä kulueriä (Nieminen & Nieminen, 2013, s. 208). Investointien rahoitusrakennepäätökset vaikuttavat laitoksen pääomakustannuksiin ja velanottoon. Niiden optimointi on keskeinen tehtävä vesihuoltolaitoksen taloushallinnossa. Saneeraus- ja uusinvestointeja käsitellään erikseen, koska niiden rahoitusperiaatteet voivat poiketa toisistaan, ja ne esitetään myös visualisoinneissa omina luokkina.

Investointisuunnitelmassa jokainen investointi jaetaan kahteen rahoituskomponenttiin: omarahoitukseen ja rahoituslainaan. Jakoa varten käyttäjä voi valita kahden rahoitusmallin välillä, kiinteä ja portaittainen malli. Kiinteässä mallissa määritellään kynnyksarvo ja kiinteä omavastuussumma siten, että alle kynnyksarvon jäävät investoinnit rahoitetaan kokonaan omalla pääomalla. Kynnyksarvon ylittävistä investoinneista omarahoitusosuus pidetään vakiona ja ylittävä osa katetaan lainalla. Portaittaisessa mallissa lainoitusosuus kasvaa investoinnin koon mukana porrastetusti, jolloin pienet investoinnit rahoitetaan pääosin omalla pääomalla ja

suuret hankkeet lainalla. Kuviossa 26 on esitetty vesihuoltolaitoksen 20-vuotisjakson investointien rahoituserittely vuositasolla.



Kuvio 26. Investointien rahoituserittely erään investointiskenaarion mukaan case-vesihuoltolaitoksella.

5.4.1 Investointien rahoitus

Vesihuoltolaitoksen rahoituksen suunnittelu lähtee investointitarpeen tunnistamisesta. Ilman kuvaa saneeraustarpeesta ja uusinvestoinneista ei voida arvioida vieraan pääoman määrää tai aikataulua. Lainarahoitus on keskeinen osa pääomarakennetta, koska vesihuolto on pääomavaltaista. Verkostot rakennetaan vuosikymmeniksi, mutta investoinnit on rahoitettava lyhyemmällä aikavälillä.

Velkaehtoien valinta on strateginen päätös. Korkeus, takaisinmaksuaika ja lyhennysrakenteen vaikuttavat suoraan vuotuisen vapaaseen kassavirtaan ja velkaantumistasoon kehitykseen, mikä määrittää laitoksen kykyä ottaa vastaan uusia rahoitustarpeita. Mallissa lainojen takaisinmaksu perustuu yhden kynnyksen kaksiluokkaiseen lyhennysmalliin. Pienemmille ja kynnyksen ylittävillä lainoilla sovelletaan eri lyhennystasojia. Vuotuinen lyhennys pidetään kiinteänä, ja jaollisuudesta jäävä jäännöserä maksetaan viimeisenä eränä. Lyhennys alkaa investoinnin toteutusvuonna yhdessä pääomakustannusten kanssa. Etuna on, että pitkän aikavälin lyhennystarvetta voidaan arvioida ilman sopimuskohtaista mallinnusta.

Vuotuinen korkokustannus lasketaan vuoden alun jäljellä olevasta pääomasta. Lainat ovat tyypillisesti vaihtuvakorkoisia, jolloin korko muodostuu Euribor 12 kk -viitekorosta ja lainakohtaisesta marginaalista. Uusille rahoituslainoille sovelletaan oletuksena kiinteää korkoa samalla marginaalilla. Euribor-aikasarja perustuu historialliselta osin Euroopan keskuspankin (EKP) toteutuneisiin korkoihin ja tulevaisuuden osalta ennustemoduulin vuosikeskiarvoihin. Korkoennuste kytkeään inflaation kehitykseen, koska EKP:n ohjaukorkopäätökset tehdään osin inflaation pohjalta. Lähtötietona on Tilastokeskuksen kuluttajahintaindeksin N-BEATS-ennusteen vuosimuutos, josta korko johdetaan inflaatioon skaalautuvasta peruskorosta, marginaalista sekä logaritmisesta kiihtyvyytermistä. Liitteessä 5 esitetään korkohistoria ja Euribor 12 kk -viitekoron pitkän aikavälin estimointiennuste.

Estimointimallissa Euribor-korko muodostetaan kolmesta perusosasta: inflaation tasosta, inflaatioon skaalautuvasta spreadista eli koron ja inflaation välisestä erosta sekä kahdesta epälineaarista kiihtyvyysskorjauksesta. Yläpään korjaus aktivoituu inflaation ylittäessä EKP:n 2 %:n tavoitteen, kun taas alapään korjaus toimii deflaatiosuojana inflaation laskeessa alle 0,5 %. Mallin syötemuuttuja π on kuluttajahintaindeksin kuukausittaisen vuosimuutoksen taaksepäin liukuva keskiarvo

$$\pi(t) = \frac{1}{N} \sum_{i=0}^{N-1} \pi_{KHI}(t-i), \quad (94)$$

missä N on liukuvan keskiarvon tasoitusjakso. Tasoitus vaimentaa satunnaista kuukausivaihtelua, jotta malli reagoi vain kestäviin muutoksiin. Koska tasoitus käyttää vain mennyttä dataa, sama määrittely toimii sekä historiallisessa validoinnissa että ennustamisessa.

Korko r määräytyy yhtälöllä

$$r(\pi) = \max(0, \pi) + s \times \frac{\pi}{\pi^*} + K_{\text{ylä}}(\pi) + K_{\text{ala}}(\pi), \quad (95)$$

missä π^* on Euroopan keskuspankin inflaatiotavoite 2 % ja s maturiteettikohtainen spread. Kiihtyvyysskorjaukset K_i määritellään seuraavilla kaavoilla

$$K_{\text{ylä}}(\pi) = \begin{cases} s \times \frac{\pi}{\pi^*} \alpha \ln\left(\frac{\pi}{\pi^*}\right), & \pi > \pi^* \\ 0, & \text{muulloin} \end{cases}, \quad (96)$$

$$K_{\text{ala}}(\pi) = \begin{cases} -s \times \alpha_{\text{ala}} \ln\left(\frac{\pi_{\text{alaraja}}}{\max(\pi, \varepsilon)}\right), & \pi < \pi_{\text{alaraja}}, \\ 0, & \text{muulloin} \end{cases} \quad (97)$$

missä π_{alaraja} 0,5 % ja ε on 0,1 %, joka varmistaa korjauksen jatkuvuuden negatiivisilla inflaatioarvoilla. Parametri α säätelee yläpään epälineaarisuutta, koska negatiivinen arvo vaimentaa korkojen nousua kansantalouden korkeassa inflaatiossa, positiivinen voimistaa sitä. Vastaavasti positiivinen α_{ala} painaa korkotasoa alaspäin deflaatiouhan kasvaessa, negatiivinen nostaa sitä. Välialueella $0,5 \% \leq \pi \leq 2 \%$ korjaustermit ovat nollija ja malli palautuu lineaariseen muotoon

$$r(\pi) = \max(0, \pi) + s \times \frac{\pi}{\pi^*}. \quad (98)$$

Kertoimet α ja α_{ala} estimoidaan suljetun muodon pienimmän neliösumman menetelmällä omilla aktivointialueillaan, jolloin kumpikin parametri sovitetaan vain niillä havainnoilla, joilla vastaava korjaustermi on aktiivinen. Mallin suorituskyky arvioidaan koko aineistolla tunnusluvuilla R^2 , RMSE ja MAE.

5.4.2 Investointien poistosuunnitelma

Investointien poistolaskenta perustuu kirjanpitolain 1336/1997 5 luvun 5 §:ään, jonka mukaan pysyviin vastaaviin kuuluvan aineellisen omaisuuden hankintameno aktivoidaan ja kirjataan vaikutusaikanaan suunnitelman mukaan poistoina kuluksi. Poistosuunnitelmaa tulee arvioida säännöllisesti tilinpäätöstä laadittaessa (Kirjanpitolautakunta, 2007, s. 14–15). Kunnallisissa vesihuoltoyhtiöissä poistosuunnitelman merkitys on erityisen suuri, sillä käyttöomaisuuden kirjanpitoarvo muodostaa keskeisen osan taseen vastaavia ja vaikuttaa suoraan yhtiön pääomakustannuksiin.

Omaisuuksienhallintamallissa kullekin käyttöomaisuuden hyödykkeelle on annettu yksilöivä hyödyketunnus ID Hyödyke, jonka avulla hyödykkeet kytketään laskentakategorioihin. Kategoriakohtainen ryhmittely mahdollistaa poistosuunnitelman tarkastelun yksittäisen hyödykkeen tasolla sekä toiminnallisina kokonaisuuksina. Kategoriat on jaoteltu vesi- ja jätevesitoiminnoittain ja kattavat vesihuollon keskeisimmät käyttöomaisuuserät, kuten rakennukset, vesijohto- ja viemäriverkos-

tot, laitteistot, koneistot, irtaimiston sekä autokaluston. Luokittelu noudattaa kirjanpitoasetuksen 1339/1997 1:6 §:n mukaista aineellisten hyödykkeiden ryhmitelyä rakennuksiin ja rakennelmiin, koneisiin ja kalustoon sekä muihin aineellisiin hyödykkeisiin.

OmaisuuDENhallintamalli tukee kahta kirjanpitolain hyväksymää poistomenetelmää, tasapoisto- ja menojäännöspoistomenetelmää. Tasapoistomenetelmässä vuotuinen poisto on vakio koko pitoajan. Poistettava summa lasketaan jakamalla poistopohja tasaisesti pitoajalle

$$D = \frac{C - R}{n}, \quad (99)$$

missä D on vuotuinen tasapoisto, C on hyödykkeen hankintameno, R on ennakoitu jäännösarvo ja n on taloudellinen pitoaika vuosina. Tasapoistomenetelmä soveltuu hyödykkeille, joiden taloudellinen kuluminen on likimain lineaarista ajassa.

Menojäännöspoistomenetelmässä vuotuinen poisto määräytyy hyödykekohtaisena suhteellisena osuutena kirjanpitoarvon poistamattomasta osasta. Tilikauden t poisto D_t lasketaan kaavalla

$$D_t = p \times B_t^* = p(B_t + I_t), \quad (100)$$

missä p on poistoprosentti, B_t^* on poiston laskentaperuste tilikauden t lopussa ennen poistoa, B_t on hyödykkeen kirjanpitoarvo tilikauden t alussa ja I_t on tilikauden t aikana tehdyt investointilisäykset. Tilikauden t lopun ja seuraavan tilikauden $t + 1$ alun kirjanpitoarvo saadaan vähentämällä poisto laskentaperusteesta

$$B_t = B_t^* - D_t. \quad (101)$$

OmaisuuDENhallintamallissa kirjanpitoarvo päivitetään vuosittain kaavalla

$$B_{t+1} = (B_t + I_t) \times (1 - p). \quad (102)$$

Investointilisäykset otetaan poistosuunnitelmaan siitä vuodesta alkaen, jolloin omaisuserä aktivoidaan taseeseen. Suomen Kuntaliiton (2019, s. 23–27) ohjeistuksen mukaan ennen käyttöönottoa olevat investoinnit käsitellään keskenräisenä rakentamisena eivätkä kerrytä poistoja. Periaate on ohjelmistossa toteutettu kahdella erillisellä arvolla. Aktivoitu investointiarvo sisältää kirjanpitoon jo sisällytetyt investoinnit, joista poistojen laskenta alkaa. Investointien tuleva arvo

puolestaan kuvaa vielä aktivoimatonta, suunniteltuja investointeja, joista poistoja ei vielä synny.

Investointilisäykset kohdennetaan ID-hyödykekohtaisesti, ja samalle ID:lle useana rivinä kirjatut investoinnit summataan yhteen kullekin vuodelle. Aggregointi mahdollistaa useiden samanaikaisten investointihankkeiden, kuten rinnakkaisten verkostonosien saneerausten, yhdistämisen yhtenäiseen poistolaskentaan ilman poistotilikohtaista seuranta. Poistosuunnitelman laskenta-aikaväli kattaa 23 vuotta. Se alkaa kaksi vuotta ennen kuluva vuotta ja päättyy 20 vuotta sen jälkeen. Pitkän aikavälin tarkastelu on välttämätöntä investointioptimoinnille, koska poistot jakautuvat vuosikymmenten ajalle ja vaikuttavat pitkäkestoisesti vesihuoltolaitoksen omarahoitusosuuden kehitykseen ja velanoton tarpeeseen.

5.5 Koneopetettu taloussuunnitelma

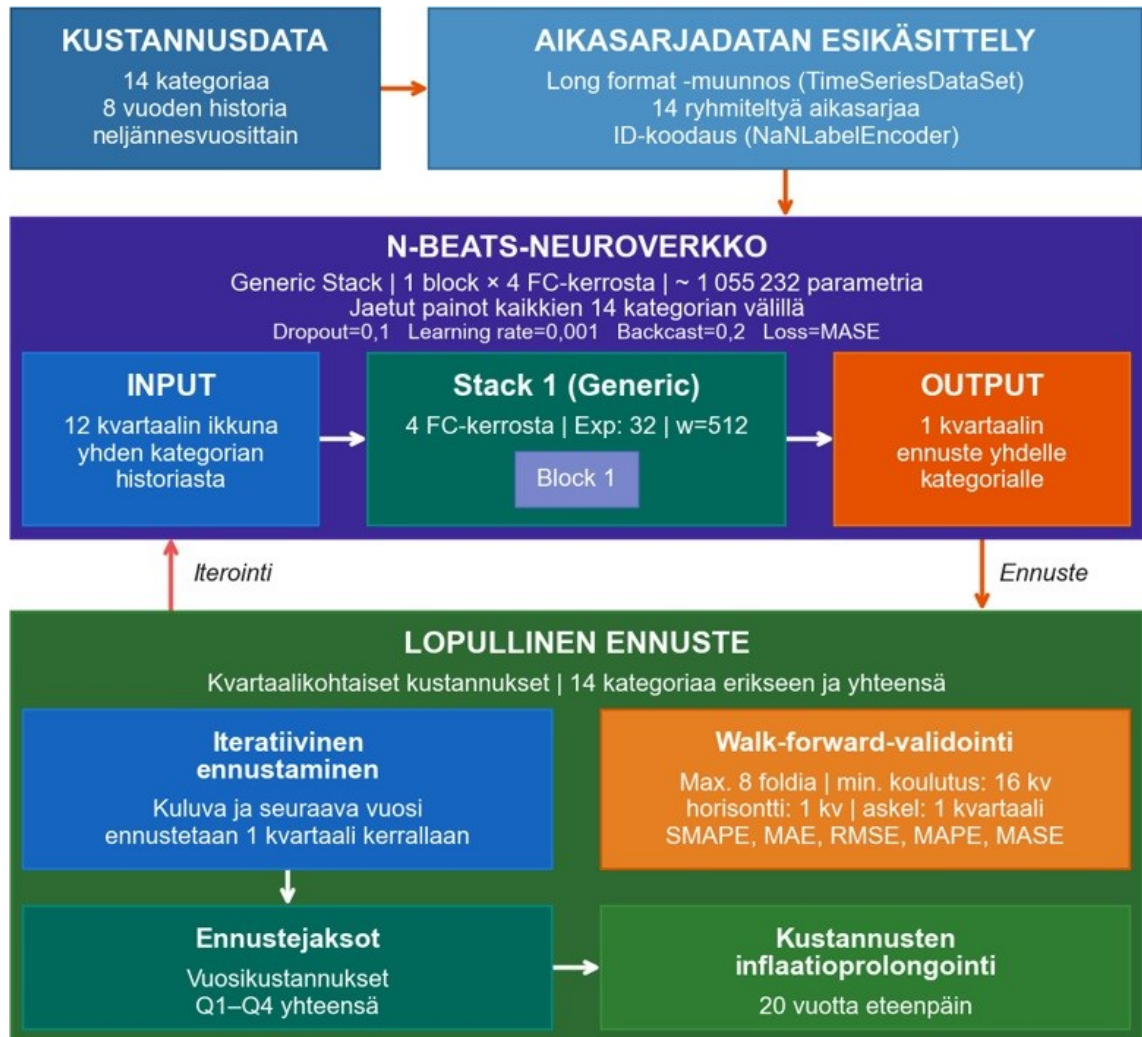
5.5.1 Operatiivisten kustannusten ennustaminen

Vesihuoltolaitoksen operatiivisten kustannusten ennustemalli pohjautuu N-BEATS-arkkitehtuurin Generic-muotoon. Malli on yksimuuttujainen, eli kukin eteenpäinpäästö ottaa syötteekseen yhden kustannuskategorian historiallisen ikkunan ja tuottaa ennusteen samalle kategorialle. Neljätoista operatiivista kustannuskategoriaa kuvataan erillisinä aikasarjoina, jotka jakavat saman verkon parametrit kategoriatunnukseen perustuvan ryhmittelyn kautta.

Täysin yhdistetyt kerrokset operoivat yhden sarjan ikkunalla, eikä kategorioiden välistä suoraa tiedonsiirtoa esiinny yksittäisen ennustepäästön aikana. Riippuvuudet välittyvät koulutusprosessin kautta, kun kaikkien neljäntoista sarjan otokset päivittävät samoja painoja. Lähestymistapa vastaa Oreshkinin ym. (2020, s. 8) esittämää monitehtäväasetelmaa, jossa muista sarjoista kertyneet säännönmukaisuudet tukevat yhden sarjan ennustetta jaettujen parametrien välityksellä.

Kuvio 27 havainnollistaa mallin tietovirran neljäntoista kategorian kahdeksan vuoden kvartaalihistoriasta lopulliseen ennusteeseen. Havainnot muunnetaan long-

formaattiin ja ryhmitellään neljäksitoista aikasarjaksi. N-BEATS-verkko tuottaa yhden kvartaalin ennusteen kahdentoista kvartaalin eli kolmen vuoden kontekstista. Iteratiivisessa prosessissa tuotettu arvo liitetään takaisin kontekstiin, jolloin saadaan ennuste kuluvan vuoden loppuun ja koko seuraavalle vuodelle. Yleistyskykyä arvioidaan laajenevan ikkunan walk-forward-validoinnilla. Kvartaalienusteet kootaan vuosikustannuksiksi ja prolongoidaan inflaatiokorjattuna 20 vuotta eteenpäin.

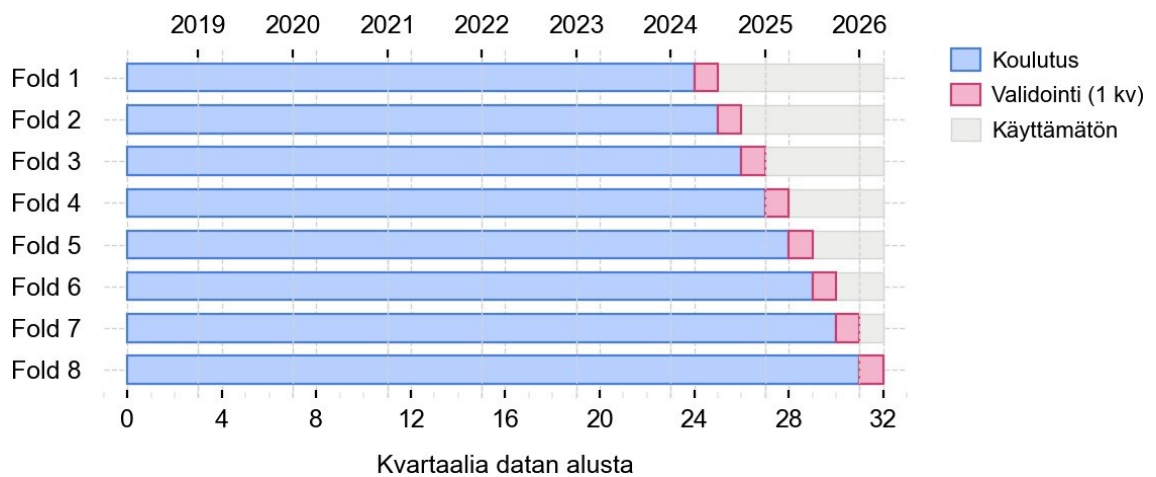


Kuvio 27. N-BEATS-kustannusennustemallin arkkitehtuuri.

Neuroverkko koostuu yhdestä Generic Stack -pinosta, jossa on yksi blokki. Blokin runkona toimii neljä peräkkäistä täysin yhdistettyä kerrosta, joiden leveys on 512 neuronialla ja joissa käytetään ReLU-aktivoitua sekä kymmenen prosentin dropoutia. Rungon päästä haarautuu kaksi rinnakkaista projektiota, jotka tuottavat 32-ulotteiset theta-vektorit backcast- ja forecast-haaroille. Theta-vektorit kerrotaan opituilla kantafunktiolla, joista syntyy kahdentoista kvartaalin konteksti ja

ennuste seuraavalle kvartaalille. Ilman analyttisiä rajoitteita sama pino mallintaa trendin ja kausivaihtelun ilman erillistä komponenttijakoa.

Walk-forward-validointi toteutetaan laajenevan ikkunan periaatteella kahdeksassa peräkkäisessä foldissa siten, että koulutusjakso kasvaa yhden kvartaalin verran foldia kohti ja seuraava kvartaali toimii validointihorisonttina (kuvio 28). Ensimmäisessäkin foldissa koulutusjakso kattaa kuusitoista kvartaalia eli neljä vuotta. Aikajärjestyksen säilyminen estää tulevaisuuden tiedon vuotamisen koulutusjoukkoon, jolloin arvio vastaa tuotantotilannetta, jossa malli koulutetaan uudelleen uuden kvartaalihavainnon myötä. Menettely tuottaa kahdeksan riippumattonta ennustevirhettä, joiden keskiarvo ja vaihtelu kuvaavat yleistyskykyä.



Kuvio 28. Kustannusennustemallin walk-forward-validoinnin periaate laajenevalla ikkunalla.

Liitteessä 6 esitetään mallin suorituskyvyn analyysi kolmessa kuvassa. Ensimmäinen kokooa walk-forward-validoinnin virhemittarien keskiarvot kahdeksalta foldilta (SMAPE 60,20 %, MAE 5 500 €, RMSE 9 900 €, MAPE 38,34 %, MASE 0,810). Toinen näyttää loss-arvot foldeittain sekä niiden kehityksen viimeisen foldin koulutuksen aikana. Kolmas esittää gradienttivirran kerroksittain viimeisellä epookilla ja kymmenen suurimman gradientin kehityksen koko koulutuksen ajan.

5.5.2 Pitkän aikavälin liikevaihtotavoitteen asettaminen

N-BEATS-mallin tuottamat kustannusennusteet kattavat enintään kahdeksan kvartaalin horisontin. Pitkän aikavälin taloussuunnittelun vaatimukset edellyttävät

lisäksi laskentaa, jolla kustannusten kasvua voidaan johtaa inflaatio-odotteiden perusteella. Malli hyödyntää korkoa korolle -periaatetta, jonka koronlaskun kaavan Niskanen ja Niskanen (2020, s. 85) määrittelevät yleisesti koskemaan periodillista aikajaksoa. Kaava voidaan esittää muodossa

$$FV_n = PV_c \times \prod_{t=1}^n (1 + r_t), \quad (103)$$

missä PV_c on lähtövuoden kustannukset, jonka N-BEATS-malli ennustaa, FV_n on n vuoden kuluttua ennustetut kustannukset, r_t on vuoden t inflaatioprosentti. Inflaatio-odotteet johdetaan kuluttajahintaindeksin aiemmista havainnoista ja N-BEATS-mallilla ennustetuista vuosimuutosprosentteista.

Kustannusnäkökulmainen liikevaihtotavoite LVT johdetaan kassavirtalaskelmasta, jossa liikevaihdon tulee kattaa operatiiviset kustannukset K , omarahoitteiset investoinnit I , lainojen lyhennykset L , rahoituskulut R sekä tuloverot T , ja tämän lisäksi tuottaa kassavirtaperusteinen voittomarginaalitavoitteen w osuus liikevaihdosta. Voittomarginaalitavoite sisältää sekä omistajien tuottovaatimuksen että puskurin kassan vahvistamiseen. Nieminen ja Nieminen (2019, s. 401) korostavat, että kassabudjetoinnissa käyttöpääomaeriin liittyvät toimenpiteet vaativat tuekseen tarkkoja ennusteita tulevista kassavirroista sekä maksuehtojen muuttamisessa päätöksenteon tueksi tarvitaan tietoja kassavirtojen ajoittumisesta ja suuruudesta. Tuloverot määritetään liikevoitosta EBIT, joka lasketaan

$$EBIT = LVT - K - P, \quad (104)$$

missä P on poistot. Verojen T osuus kassavirrasta on tällöin

$$T = v \times EBIT = v(LVT - K - P - R), \quad (105)$$

missä v on tuloveroprosentti. Liikevaihtotavoite saadaan ratkaisemalla kassavirtaehto

$$LVT = K + I + L + R + v(LVT - K - P - R) + w \times LVT \quad (106)$$

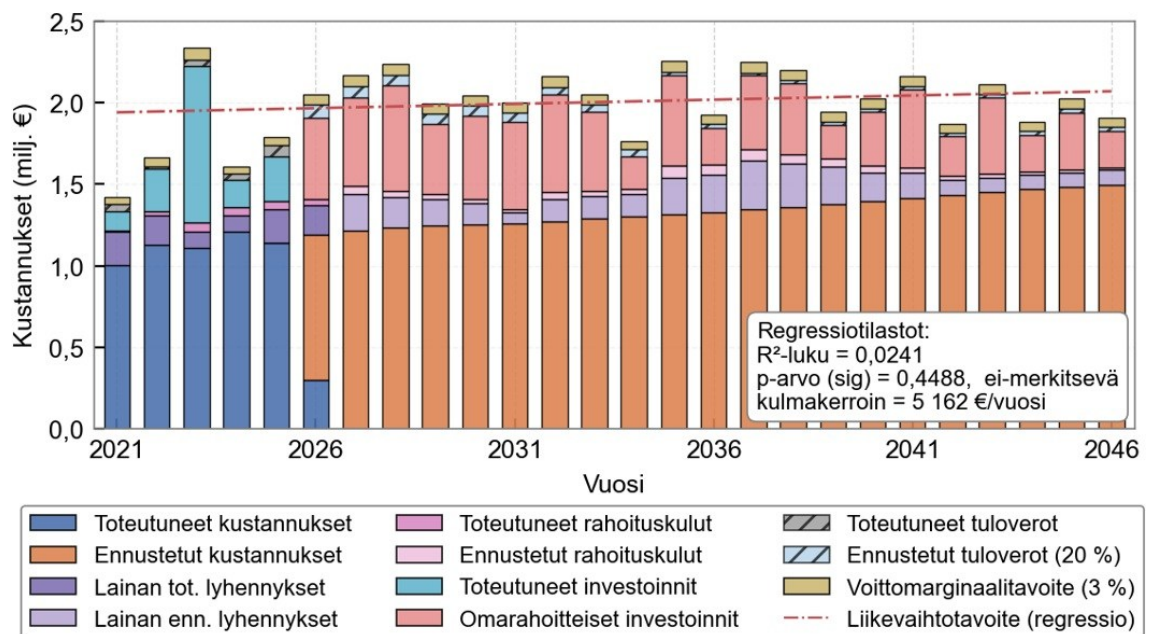
muuttujan LVT suhteen, jolloin saadaan

$$LVT = \frac{K(1 - v) + I + L + R(1 - v) - P \times v}{1 - v - w}. \quad (107)$$

Poistot P eivät itsessään aiheuta kassavirtaa, koska ne sisältyvät osoittajaan termin $-P \times v$ kautta, koska poistot toimivat verokilpenä ja siten pienentävät kassasta lähtevien tuloverojen määrää. Korkomaksujen verovähennysoikeudesta

johtuen velkaantumisella sanotaan olevan ns. verosuoja (Knüpfer & Puttonen, 2018, s. 195). Vastaavasti termi $R(1 - v)$ osoittajassa heijastaa rahoituskulujen verovähennyskelpoisuutta, koska rahoituskulujen kassavirtarasitteesta vähennetään verokilven arvo $v \times R$. Tällöin valtio rahoittaa osan rahoituskuluista alennettujen verojen kautta. Laskennassa ei huomioida satunnaisia eriä eikä rahoitustuottoja, kuten korkotuottoja kassavaroista tai osinkotuottoja sijoituksista, jota opinnäytetyössä perustellaan yksinkertaisemmalla laskennalla. Tämän rajauksen seurauksena malli antaa hieman varovaisemman arvion liikevaihtotavoitteesta, kuin täydellinen tuloslaskelmaperusteinen tarkastelu antaisi.

Kuvio 29 esittää liikevaihtotavoitteen muodostumisen kassavirtanäkökulmasta. Pinotut pylväät koostuvat seitsemästä komponentista, jotka on jaettu toteutuneisiin ja ennustettuihin eriin: operatiiviset kustannukset, lainojen lyhennykset, rahoituskulut, investoinnit, tuloverot sekä voittomarginaalitavoite.

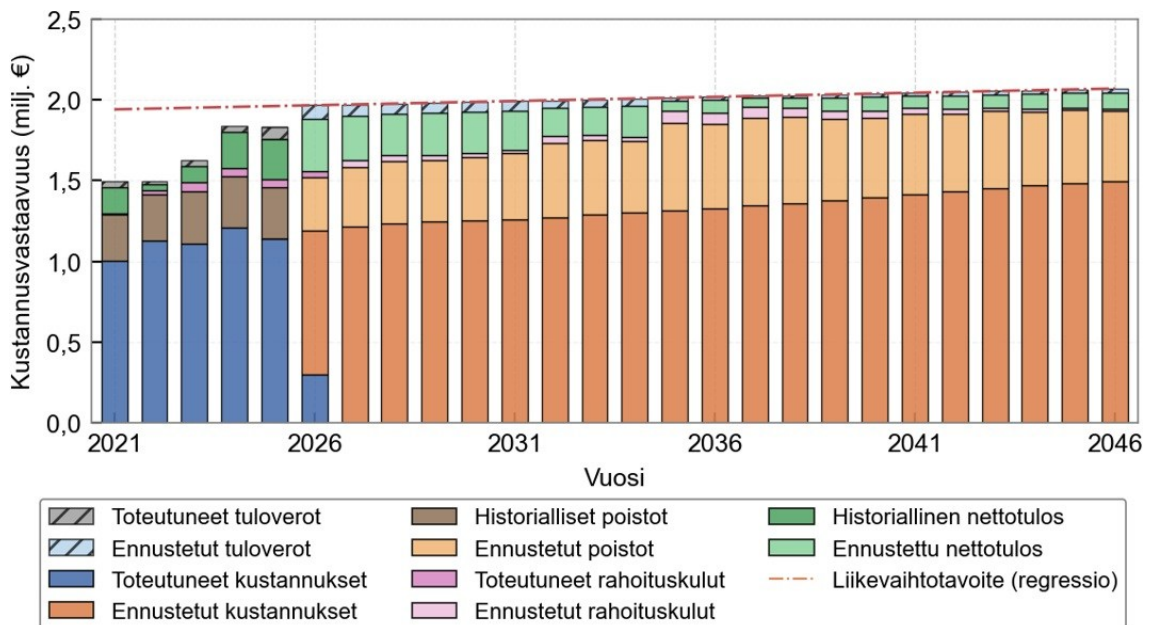


Kuvio 29. Liikevaihtotavoitteen muodostuminen kassavirtanäkökulmasta case-vesihuoltolaitoksessa erään skenaarion mukaisesti.

Tuloverot lasketaan ennustevuosille tuloveroprosentilla liikevoitosta ja voittomarginaalitavoite kolmen prosentin osuutena liikevaihdosta. Punainen katkoviiva kuvaa pylväiden kokonaiskorkeudesta laskettua lineaarista regressiota, joka ilmaisee pitkän aikavälin liikevaihtotavoitteen kehityssuunnan. Regressiosuora nou-

see loivasti, mikä heijastaa kustannusten inflaatiopohjaista kasvua pitkällä aikavälillä. Yksittäisten vuosien pylväskorkeudet vaihtelevat regressiosuoran ympärillä, ja poikkeamat selittyvät pääosin investointien ja lainanlyhennysten ajoituksesta.

Kuvio 30 esittää saman ajanjakson liikevaihdon hajotettuna tuloslaskelman peruskaavan mukaisiin komponentteihin. Kirjanpitokäytännön mukainen tuloslaskelma rakentuu hierarkkisesti, koska liikevaihdosta vähennetään ensin operatiiviset kulut, sitten poistot, jolloin muodostuu liikevoitto EBIT, ja tästä edelleen rahoitustuotot ja -kulut, jolloin saadaan tulos ennen satunnaisia eriä. Tuloksesta ennen tilinpäätössiirtoja ja veroja lasketaan tuloverot, joiden jälkeen jää tilikauden tulos eli nettotulos (Niskanen & Niskanen, 2020, s. 45–48).



Kuvio 30. Liikevaihdon kehitys tuloslaskelman näkökulmasta case-vesihuoltolaitoksessa erään skenaarion mukaisesti.

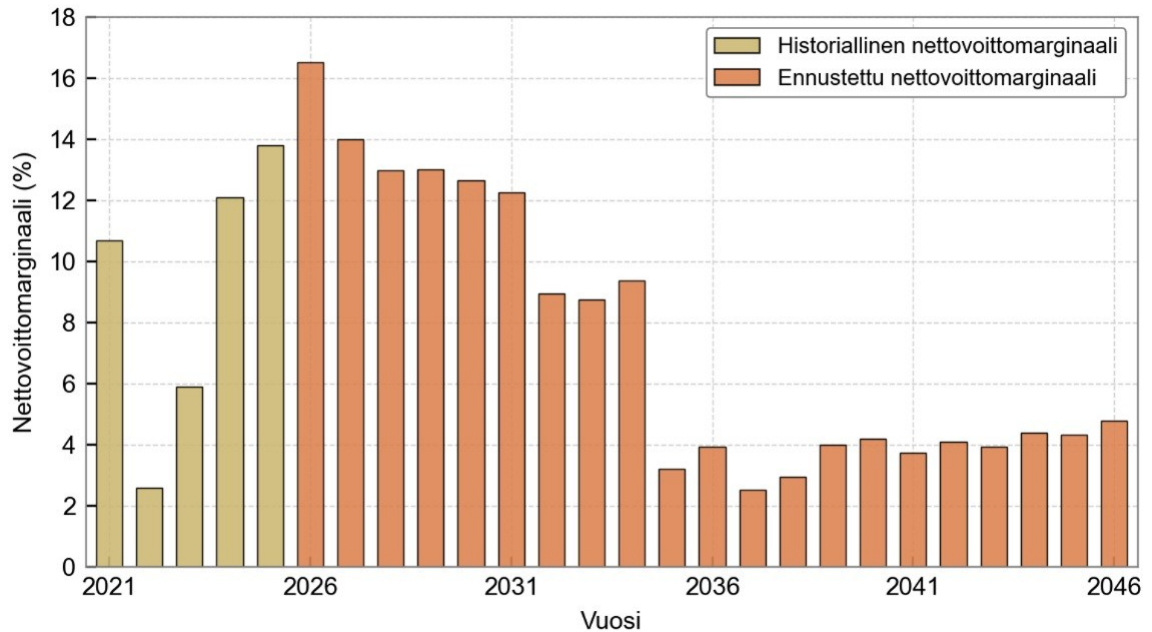
Liikevaihtotavoite perustuu vesihuoltolaitoksen rahaprosessiin

$$LVT = K + P + R + \text{Nettotulos} + \text{Verot} \quad (108)$$

ja se voidaan tulkita pylvään viisikerroksisena rakenteena alhaalta ylös: operatiiviset kustannukset, poistot, rahoituskulut, nettotulos ja tuloverot. Kerroksista voidaan lukea EBIT kaavalla (104), joka näkyy poistokerroksen yläreunan ja pylvään huipun välisenä korkeutena sekä

$$\text{Tulos ennen satunnaisia eriä} = \text{EBIT} - R, \quad (109)$$

joka näkyy rahoituskulukerroksen yläreunan ja pylvään huipun välisenä korkeutena. Verot lasketaan voitosta rahoituserien jälkeen $v(\text{EBIT} - R)$, jolloin nettotulos on $(1 - v)(\text{EBIT} - R)$. Pylvään ylin osa jakautuu siten nettotulokseen ja verokerrokseen vakiosuhteessa $(1 - v):v$ eli 80:20, kun v on 20 prosenttia. Kuvio 31 esittää saman ajanjakson nettovoittomarginaalin prosenttipohjaisena tunnuslukuna.



Kuvio 31. Nettovoittomarginaalin ennustettu kehitys case-vesihuoltolaitoksella.

5.5.3 Vesihuollon maksujen automaattinen hinnoittelu

Vesihuolto on yhteiskunnallisesti välttämätön infrastruktuuri, jonka rahoitus perustuu asiakkailta perittäviin maksuihin. Vesihuoltolaki (119/2001) velvoittaa vesihuoltolaitokset asettamaan maksut kustannusvastaaviksi, kohtuullisiksi ja tasa-
puolisiksi. Maksujen on myös turvattava laitoksen pitkän aikavälin toimintakyky sekä investointien rahoitus. Käytännössä tämä tarkoittaa, että kunkin asiakasryhmän tulee kattaa omasta palvelustaan aiheutuvat kustannukset eikä mitään asiakasryhmää saa perusteettomasti tukea toisen kustannuksella. Kustannusvastaavuuden ohella hinnoittelussa huomioidaan useita muita tavoitteita, joita ovat tulojen riittävyys, maksutuottojen vakaus, hintojen ennustettavuus ja veden käytön

tehokkuuden edistäminen. Nämä tavoitteet heijastuvat suoraan maksujen rakenteeseen ja suunnitteluperiaatteisiin.

Vesihuoltolaitoksen kustannusrakenne muodostuu suurelta osin kiinteistä kustannuksista, kuten verkoston ja laitosten pääoma- ja ylläpitokuluista, jotka eivät riipu lyhyen aikavälin talousveden kulutusvaihteluista. Tämän vuoksi on perusteltua kerätä merkittävä osa tuloista kiinteillä perusmaksuilla, jotta laitoksen talous pysyy vakaana vedenkulutuksen muuttuessa. Käytännössä perusmaksujen osuuden kasvattaminen kokonaistuotoista onkin yksi yleisimmistä tavoista vakauttaa vesihuoltolaitoksen tulovirtaa (American Water Works Association, 2017, s. 149–150).

Kiinteiden maksujen tuottamat tulot ovat ennustettavampia kuin käyttömaksujen tuottamat tulot. Toisaalta perusmaksupainotteinen hinnoittelu heikentää veden säästön hintasignaalia ja voi vaikuttaa pienkuluttajien kokemaan hinnan kohtuullisuuteen. Perus- ja käyttömaksujen keskinäinen suhde on siten aina kompromissi taloudellisen vakauden ja vedenkulutuksen hallinnan tavoitteiden välillä.

Vesihuoltomaksujen automaattinen hinnoittelu perustuu siihen, että koko hinnasto johdetaan laskennallisesti muutamasta strategisesta lähtöarvosta sen sijaan, että yksittäiset maksut määritettäisiin käsin. Lähtötietoina toimivat liikevaihdon kokonaistavoite, kulutusennuste, kiinteän tuoton osuus sekä perusmaksun tavoiteosuus kokonaistuotosta. Näistä malli ratkaisee ensin käyttömaksun yksikköhinnan ja omakotitalon perusmaksun suuruuden. Tämän jälkeen porrastuskaavat tuottavat huoneistoluvun mukaan skaalatut perusmaksut rivi- ja kerrostaloille, ja ekvivalenssikertoimet yhdenmukaistavat mittarikokoon perustuvan hinnoittelun palvelu- ja teollisuusrakennuksille.

Ikaalisten Vesi Oy:n asiakasmaksut muodostuvat käyttö-, perus- ja liittymismaksuista. Käyttömaksu on sidottu kulutettuun vesimäärään. Perusmaksu on kiinteä kuukausimaksu, joka kattaa palveluvalmiuden pääoma- ja ylläpitokustannukset kulutuksesta riippumatta. Liittymismaksu peritään kertamaksuna uuden liittymän yhteydessä investointikustannusten kattamiseksi (Ikaalisten Vesi Oy, 2026, s. 1). Maksujen kohdentamisen tulee perustua aiheuttamisperiaatteeseen (American Water Works Association, 2017, s. 89–92).

Perusmaksuhinnoittelu perustuu joko vesimittarin nimelliskokoon tai kiinteistön asuinhuoneistojen lukumäärään. Näiden logiikoiden yhteensovittaminen on tarpeen, kun toiminta-alueella palvellaan sekä asuinrakennuksia että muita rakennusluokkia. Koska rakennusluokitus ja huoneistomäärä vaikuttavat potentiaaliin vedenkulutukseen, rivi- ja kerrostalon perusmaksu on asukasta kohden edullisempi kuin omakotitalon. Suuremmissa rakennuksissa liittymä- ja verkostokustannukset jakautuvat useamman käyttäjän kesken, jolloin huoneistokohtainen perusmaksu pienenee (Ikaalisten Vesi Oy, 2026, s. 1).

Porrastuksen toteuttamiseksi määritellään matemaattinen malli käyttöpaikan perusmaksulle

$$P_{\text{vesi}} = V_A \times C^{1+\log_6 X_A}, \quad (110)$$

missä C on asuinhuoneistojen lukumäärä, V_A on omakotitalon veden arvonlisäverollinen perusmaksu kuukaudessa ja $X_A = 0,50$ on logaritmisesta skaalauskaiväyrän muotoa ohjaava kerroin. Eksponentti $1 + \log_6(0,5) \approx 0,613$ on pienempi kuin 1, joten perusmaksu huoneistoa kohden pienenee rakennuskokoon kasvaessa. Vastaava eksponentti jätevedelle on $1 + \log_7(0,6) \approx 0,737$.

Huoneistokohtainen maksu saadaan jakamalla kokonaismaksu huoneistomäärällä

$$V_{\text{huoneisto}}(C) = \frac{V_{\text{perus}}(C)}{C} = V_A \times C^{\log_6 X_A}. \quad (111)$$

Koska $\log_6 X_A < 0$, funktio on aidosti laskeva, mikä toteuttaa hinnaston periaatteen rivi- ja kerrostalojen edullisemmasta huoneistokohtaisesta hinnasta (Ikaalisten Vesi Oy, 2026, s. 1). Jotta mittarikokoon perustuva hinnoittelu saadaan yhteismittaliseksi huoneistopohjaisen hinnoittelun kanssa, määritellään ekvivalenttihuoneistoluku $C_{\text{ekv}}(Q_n)$. Se kuvaa tietyn mittarikokoon palvelukapasiteettia asuinhuoneistoyksiköissä ja on määritelty siten, että mittariperusteinen kokonaismaksu $V^{\text{mittari}}(Q_n)$ ja huoneistoperusteinen kokonaismaksu $V_{\text{perus}}(C_{\text{ekv}})$ ovat yhtä suuret

$$V^{\text{mittari}}(Q_n) = V_A \times C_{\text{ekv}}(Q_n)^{1+b}. \quad (112)$$

Vertailukohtana käytetään mittarikokoa DN 20, jolle asetetaan ehto $C_{\text{ekv}}(Q_n, \text{DN}20) = 1$. AWWA (2018, s. 152, 365) toteaa, että ekvivalenssikertoimien johtamisessa

tulee huomioida sekä mittarin hydraulinen kapasiteetti että kyseisen asiakasluokan tyypillinen kysyntäprofiili. Jos kalibrointi halutaan sitoa suoraan nimellisvirtaamaan, voidaan käyttää kaavaa

$$C_{\text{ekv}}(Q_n) = \left(\frac{Q_n}{Q_{n,\text{DN20}}} \right)^\eta, \quad (113)$$

jossa parametri η määräytyy laitoksen valitsemasta kapasiteettiskaalausperiaatteesta

$$\eta = \frac{\ln C_{\text{ekv}}}{\ln \left(\frac{Q_n}{Q_{n,\text{DN20}}} \right)}. \quad (114)$$

Kustannusvastaavien hinnastoparametrien määrittäminen edellyttää liikevaihtotavoitteen asettamista, kulutusennusteen laatimista sekä kiinteän tuoton osuuden valintaa. Laitoksen liikevaihto R muodostuu kahdesta perusosasta

$$R = R_{\text{perus}} + R_{\text{käyttö}}. \quad (115)$$

Käyttömaksukomponentti saadaan kertomalla kulutusennuste Q_{tot} [m^3/vuosi] yksikköhinnalla p [$\text{€}/\text{m}^3$]

$$R_{\text{käyttö}} = p \times Q_{\text{tot}}. \quad (116)$$

Kiinteän tuoton osuutta kuvataan parametrilla φ , jolloin

$$R_{\text{perus}} = \varphi \times R, \quad R_{\text{käyttö}} = (1 - \varphi) \times R \quad (117)$$

Ikaalisten Vesi Oy:n (2026, s. 1) hinnaston mukaan perusmaksujen keskeisiin periaatteisiin kuuluu sekä kustannusvastaavuus että veden kulutuksen hillitseminen, mikä käytännössä asettaa φ tasolle, joka tasapainottaa nämä kaksi tavoitetta. Kun φ on valittu ja kulutusennuste tiedetään, käyttömaksun p yksikköhinta saadaan kaavalla

$$p = \frac{(1 - \varphi) \times R}{Q_{\text{tot}}}. \quad (118)$$

Vesihuoltomaksujen pitkäjänteisen kehityksen ohjaamiseksi asetetaan tavoite sille, kuinka suuri osuus vedestä ja jätevedestä saatavasta maksutuotosta tulee kertyä perusmaksuista suhteessa käyttömaksuihin. Tätä suhdetta kutsutaan perusmaksun tavoiteosuudeksi kokonaistuotosta eli PM-osuudeksi. Se määritellään talousvedelle ja jätevedelle erikseen

$$\tau_{\text{vesi}} = \frac{PM_{\text{vesi}}}{PM_{\text{vesi}} + KM_{\text{vesi}}}, \quad \tau_{\text{JV}} = \frac{PM_{\text{JV}}}{PM_{\text{JV}} + KM_{\text{JV}}}, \quad (119)$$

missä PM_{vesi} ja KM_{vesi} ovat vastaavasti veden perusmaksu- ja käyttömaksutuotot, PM_{JV} ja KM_{JV} jäteveden vastaavat tuotot. Perusmaksun tavoiteosuudeksi kokonaistuotosta määritellään strategiassa se prosenttiosuus, jonka verran vedestä ja jätevedestä saatavasta vuotuisesta maksutuotosta pyritään keräämään perusmaksuilla. Tavoitearvoiksi voidaan asettaa esimerkiksi τ_{vesi} on 40 % ja τ_{JV} 35 %. Tämä tarkoittaa, että vedenkäytön maksutuotosta 40 prosenttia kertyy kiinteistä perusmaksuista ja 60 prosenttia kulutusmääräisistä käyttömaksuista.

Tarvittavat maksukorotukset johdetaan analyttisesti PM-osuustavoitteista ja liikevaihdon kokonaistavoitteesta. Olkoon $X = PM_{\text{vesi}} + KM_{\text{vesi}}$ veden ja $Y = PM_{\text{JV}} + KM_{\text{JV}}$ jäteveden kokonaistuotto korotuksen jälkeen. PM-osuustavoitteiden suhde esitetään suhdekaavalla

$$k = \frac{X}{Y} = \frac{\tau_{\text{JV}} \times PM_{\text{vesi}}}{\tau_{\text{vesi}} \times PM_{\text{JV}}}. \quad (120)$$

Liikevaihdon kokonaistavoite L^* sidotaan X :ään ja Y :hyn ottamalla huomioon myynti muille vesihuoltolaitoksille ja muut vakiotuotot. Ratkaisemalla Y saadaan

$$Y = \frac{L^* - m \times PM_{\text{vesi}} - K_{\text{const}}}{k(1 + m \times \tau_{\text{vesi}}) + 1}, \quad (121)$$

missä m on myynnin kasvukerroin myydylle vedelle muille vesihuoltolaitoksille, K_{const} on korotuksista riippumattomat vakiotuotot ja nimittäjän termi $k(1 + m \tau_{\text{vesi}}) + 1$ kokoaa veden ja jäteveden liikevaihtokomponenttien keskinäisen sidoksen. Tästä johdetaan tarvittavat korotusprosentit

$$r_{\text{PM}} = \frac{\frac{\tau_{\text{vesi}} \times X}{PM_{\text{vesi}}} - 1}{E_{\text{PM}}}, \quad (122)$$

missä r_{PM} perusmaksun korotusprosentti ja E_{PM} on perusmaksun efektiivinen herkkyyserroin, joka muodostuu huoneistokohtaisesta kuluttajaosuudesta sekä vesimittarikokoon perustuvan hinnoittelun osuudesta, johon sovelletaan 1,2-kertaista korotuserrointa. Kun asiakaskannan huoneistolukujen ja ekvivalenttihuoneistolukujen jakaumat tunnetaan, voidaan näiden suureiden avulla määrittää käyttömaksun yksikköhinta sekä perusmaksun suuruus omakotitalolle.

5.5.4 Riskinarvio vesihuoltolaitoksen talouden tilasta

Vesihuoltolaitoksen taloudellista tilaa koskeva riskinarviointi toteutetaan DuPont-analyysiä hyödyntäen. Menetelmässä laitoksen riskiprofiilia arvioidaan kuudella osa-alueella, joista kullekin määritetään kokonaispistemäärä viisiportaisella asteikolla 1–5. Kokonaisriskitaso johdetaan osa-aluepisteiden painotettuna keskiarvona, jossa painokertoimet ovat tapauskohtaisesti konfiguroitavissa. Pisteytyksen käsitteellinen perusta nojaa taloudelliseen vakausanalyysiin, jossa vakava-raisuutta, velkarakennetta, kannattavuutta, kannattavuuden vakautta, investointiriskiä sekä konkurssiriskiä tarkastellaan rinnakkaisina ja toisiaan täydentävinä tekijöinä.

Altman (1968) kehitti Z-score-mallin yritysten konkurssiriskin ennustamiseen erotteluanalyysin avulla. Z-malli suunniteltiin julkisille teollisuusyrityksille, ja se yhdisti viisi tilinpäätöspohjaista tunnuslukua painotetuksi kokonaisindeksiksi. Altmanin (2000) mukaan päivitettyssä Z"-mallissa X_5 -muuttuja eli liikevaihto/taseen loppusumma poistettiin alkuperäisestä kaavasta, minkä jälkeen jäljelle jäävien muuttujien kertoimet estimoitii uudelleen. Z"-malli soveltuu vesihuoltolaitoksille, joiden tase on vesihuolto-omaisuuden vuoksi rakenteellisesti raskas ja pääomavaltainen, koska X_5 -muuttujan poistaminen vähentää pääomaintensiivisyys-eroista johtuvaa toimialasidonnaista vaihtelua. (Altman, 2000, s. 7–23.)

Z"-mallin kaava on (Altman, 2000, viitattu lähteessä Radivojac ym., 2021, s.14)

$$Z'' = 6,56X_1 + 3,26X_2 + 6,72X_3 + 1,05X_4, \quad (123)$$

missä osatekijät määritellään seuraavasti

$$X_1 = \frac{\text{Oma pääoma} - \text{Lyhytaikainen vieras pääoma}}{\text{Kokonaispääoma}}, \quad (124)$$

$$X_2 = \frac{\text{Nettotulos}}{\text{Kokonaispääoma}}, \quad (125)$$

$$X_3 = \frac{\text{EBIT}}{\text{Kokonaispääoma}}, \quad (126)$$

$$X_4 = \frac{\text{Oma pääoma}}{\text{Vieras pääoma}}. \quad (127)$$

Mallin tuottaman Z'' -arvon tulkinta perustuu ennalta määriteltyihin luokittelukynnysarvoihin. Taulukossa 3 esitetyt tulkintavyöhykkeet jakavat havainnot kolmeen riskiluokkaan: turvalliseen vyöhykkeeseen, välitilan vyöhykkeeseen sekä vaaralliseen vyöhykkeeseen.

Taulukko 3. Z'' -mallin tulkintavyöhykkeet (Altman, 2000, Radivojac ym., 2021, s. 14, muokattu)

Vyöhyke	Z'' -arvo	Tulkinta
Turvallinen	$Z'' > 2,60$	Matala konkurssiriski
Välitila	$1,10 < Z'' \leq 2,60$	Epävarma alue
Vaarallinen	$Z'' < 1,10$	Korkea konkurssiriski

X_1 edustaa nettokäyttöpääomaa suhteessa taseen loppusummaan ja mittaa yrityksen lyhyen aikavälin likviditeettiä. X_2 heijastaa kertyneiden voittovarojen suhdetta taseeseen ja kuvaa yrityksen kykyä rahoittaa toimintaansa sisäisesti. X_3 on kokonaispääoman tuotto liikevoittotasolla, joka on Altmanin (2000) mukaan yksittäisistä muuttujista tilastollisesti merkitsevin ennustaja. X_4 mittaa vakavaraisuutta suhteuttamalla oman pääoman vieraaseen pääomaan. (Altman, 2000, s. 7–23.) Z'' -versiossa käytetään kirjanpitoarvoja markkina-arvojen sijaan.

Omavaraisuusaste on vakavaraisuuden keskeinen tunnusluku, joka kuvaa oman pääoman suhteellista osuutta yrityksen koko pääomarahoituksesta (Knüpfer & Puttonen, 2018, s. 238–240). Tunnusluku indikoi yrityksen kykyä kestää tappioita ja rahoitusrakenteen puskurikykyä taloudellisten häiriöiden aikana. Omaisuudenhallintamallissa yksinkertaisen omavaraisuusasteen laskentakaava on

$$\text{Omavaraisuusaste} = \frac{\text{Oma pääoma}}{\text{Taseen loppusumma}} \quad (128)$$

Vesihuoltolaitoksen velkarakennetta arvioidaan omaisuudenhallintamallissa nettovelkaantumisasteella, joka mittaa korollisen vieraan pääoman määrää suhteessa omaan pääomaan sen jälkeen, kun saatavilla olevat likvidit rahavarat on vähennetty velkaa pienentävänä puskurina. Nettovelkaantumisaste lasketaan kaavalla

$$\text{Nettovelkaantumisaste} = \frac{\text{Korolliset velat} - \text{Likvidit rahavarat}}{\text{Oma pääoma}}. \quad (129)$$

Korolliset velat voivat koostua rahoituslaitoslainoista tai konserniin kuuluilta osapuolilta saaduista veloista. Tarkastelu kohdistuu korollisiin velkoihin, koska niihin liittyvät sopimusperusteiset maksuveloitteet, kuten korot ja lyhennykset, rasittavat kassavirtaa liikutuloksesta riippumatta. Likvidien rahavarojen vähentäminen osoittaa, miltä osin laitos pystyy kattamaan velkansa välittömästi omilla varoillaan (Niskanen & Niskanen, 2020, s. 59).

Ennustejakson tase-erien muodostamisessa pitkäaikaisen vieraan pääoman kehitystä mallinnetaan velkabalanssiyhtälöllä, joka päivittää vieraan pääoman määrän $VP_t^{\text{pitkä}}$ vuosittain edellisen periodin velkakannan, uusien lainanottojen ja lyhennysten avulla

$$VP_t^{\text{pitkä}} = \max(0, VP_{t-1}^{\text{pitkä}} + \text{Nostot}_t - \text{Lyhennykset}_t). \quad (130)$$

Operaattori $\max(0, \cdot)$ asettaa velkamuuttujalle epänegatiivisuusehdon ja estää siten lainakannan painumisen negatiiviseksi tilanteissa, joissa kumulatiiviset lyhennykset ylittäisivät jäljellä olevan velan. Oman pääoman kehitystä OP_t mallinnetaan vastaavasti kumulatiivisena prosessina, jossa edellisen periodin omaan pääomaan lisätään tilikauden nettotulos

$$OP_t = OP_{t-1} + \text{Nettotulos}_t. \quad (131)$$

OmaisuuDENhallintamallissa oletetaan, ettei vesihuoltolaitos jaa osinkoja eikä toteuta oman pääoman ehtoisia rahoitusjärjestelyjä ennustejakson aikana, joten kaavan (131) mukaan tilikauden nettotulos jää oman pääoman ainoaksi ajuriksi. Oletus on yksinkertaistava, mutta riskianalyysin näkökulmasta perusteltu, koska se eristää pääomarakenteen kehityksen keskeisiin tulos- ja velkavirtoihin. Vesihuoltosektorilla oletusta voidaan pitää empiirisesti uskottavana, koska kunnalliset liikelaitokset eivät tyypillisesti harjoita systemaattista voitonjakopolitiikkaa. Kokonaispääoma K_t saadaan tällöin oman ja vieraan pääoman summana

$$K_t = OP_t + VP_t. \quad (132)$$

Vesihuollossa investointien riittävyys suhteessa pääomakannan kulumiseen on keskeinen taloudellisen kestävyuden tekijä. Omaisuudenhallintamalli arvioi investointiriskiä kahdesta toisiaan täydentävästä näkökulmasta. Ensimmäinen näkökulma on kirjanpidolliseen kulumiseen perustuva investointi/poisto-suhdeluku ja toinen reaalitaloudelliseen kulumiseen perustuva kumulatiivinen saneerausvaje. Investointi/poisto-suhdeluku kuvaa, kattavatko vuotuiset investoinnit niitä vastaavan ajanjakson poistot eli pidetäänkö pääomakanta vähintäänkin kirjanpidollisen kulumisvauhtinsa tasolla. Suhdeluku lasketaan ohjelmassa kaavalla

$$\text{Investointi/poisto} = \frac{\text{Investoinnit}}{\text{Poistot}}. \quad (133)$$

Taulukko 4 esittää ohjelman riskimittarien rakenteet, painot ja pisteytysperiaatteet.

Taulukko 4. Riskimittarien pisteytysrakenne ja painot.

Osa-alue	Tunnusluku	Paino	Pisteytyslogiikka
Vakavaraisuus	Omavaraisuusaste	0,20	> 50 % → 5; > 40 % → 4; > 30 % → 3; > 20 % → 2; muuten 1
Velkarakenne	Nettovelkaantumisaste	0,15	< 0,50 → 5; < 0,80 → 4; < 1,20 → 3; < 2,00 → 2; muuten 1
Kannattavuus	ROE (ennustejakson keskiarvo)	0,20	> 12 % → 5; > 8 % → 4; > 5 % → 3; > 2 % → 2; muuten 1
Kannattavuusvakaus	ROEn keskihajonta	0,15	< 1 % → 5; < 2 % → 4; < 4 % → 3; < 6 % → 2; muuten 1
Investointiriski	Investoinnit / Poistot	0,15	> 1,20 → 5; > 1,00 → 4; > 0,80 → 3; > 0,60 → 2; muuten 1
Z-score	Altman Z''	0,15	> 2,60 → 5; > 2,10 → 4; > 1,10 → 3; > 0,50 → 2; muuten 1

Kokonaisriskipiste lasketaan painotettuna keskiarvona

$$R = w_1 p_1 + w_2 p_2 + \dots + w_6 p_6, \quad (134)$$

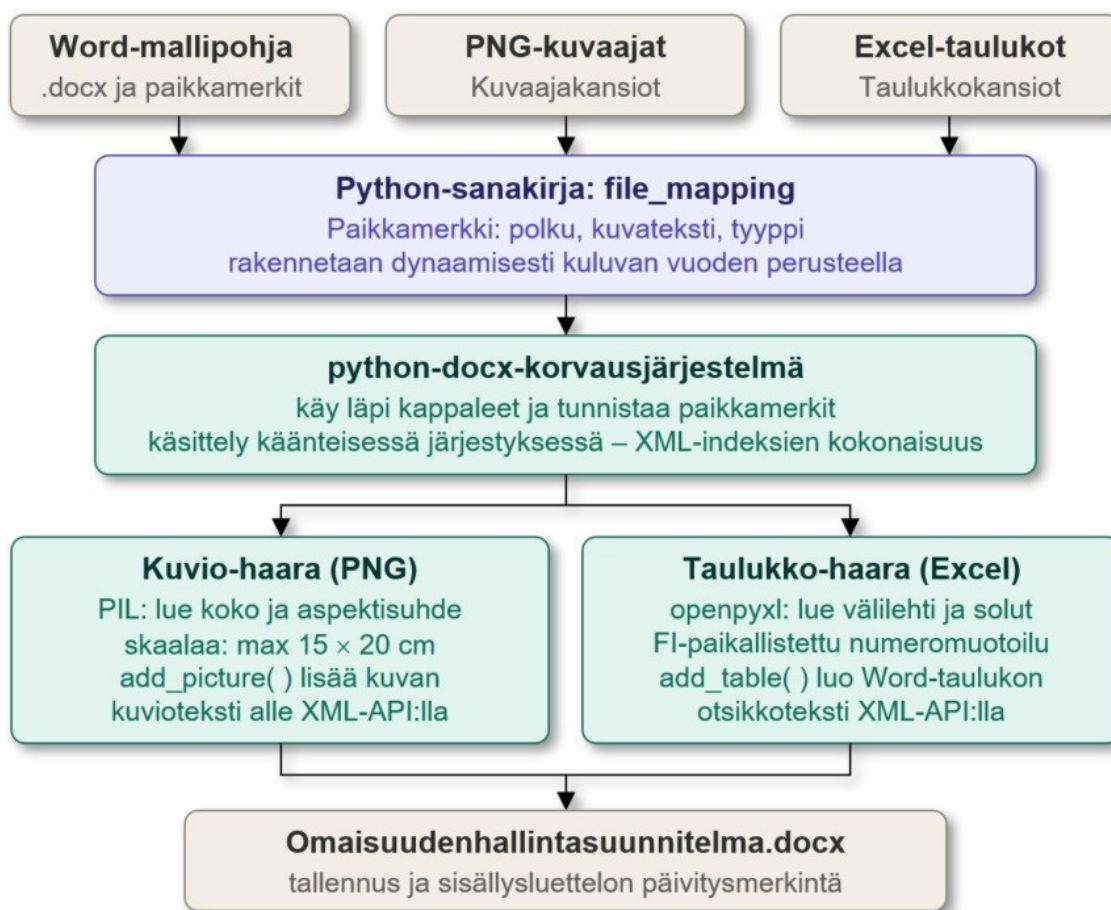
missä w_i on osa-alueen i paino kokonaissumman ollessa 1,0 ja p_i on pisteytysasteikolla 1–5. Tuloksena saatu kokonaispistemäärä luokitellaan neljään riskiluokkaan: matala riski ($R \geq 4,0$), kohtalainen riski ($3,0 \leq R < 4,0$), korkea riski ($2,0 \leq R < 3,0$) ja erittäin korkea riski ($R < 2,0$).

OmaisuuDENhallintamallin DuPont-analyysin ROE ja riskiprofiilin tulokset esitetään liitteen 7 kuvaajissa. Ensimmäinen kuvaaja erittelee oman pääoman tuoton DuPont-kaavan kolmeen osatekijään eli nettovoittomarginaaliin, pääoman kierto-nopeuteen ja rahoituskertoimeen. Toinen kokooa vesihuoltolaitoksen kokonais-riskin tutkakaavioon, jossa nykyhetkeä verrataan ennustejakson loppuun kuuden riskiulottuvuuden suhteen, ja täydentää näkymää kannattavuuden, maksuval-miusriskin ja pääomarakenteen aikasarjoilla. Kolmas kuvaaja tarkastelee Altmanin Z"-konkurssiriskimittarin kehitystä sen neljän osatekijän painotettuina vaiku-tuksina ja peilaa tasekannan kasvua inflaatiokorjattuun nimellisarvoon. Neljäs vertaa saneeraus- ja uusinvestointeja vuotuisiin poistoihin ja esittää kumulatiivi-sen saneerausvajeen kahteen teoreettiseen tarpeeseen perustuen, jotka ovat in-flaatiokorjattu nimellisarvo ja vuosipoistojen taso.

5.6 Automatisoitu omaisuudenhallintasuunnitelman dokumentointi

Vesihuoltolaitoksen omaisuudenhallintasuunnitelma rakentuu laajan tietopohjan varaan, joka kattaa inflaatioennusteet, investointi- ja kustannuslaskelmat, hinnoit-telumallit sekä riskianalyysit. Dokumentointi tuotetaan automaattisesti osana omaisuudenhallintajärjestelmää. Lähtökohtana on Microsoft Word -muotoinen mallipohja, joka määrittelee suunnitelman koko rakenteen. Mallipohjaan on sijoit-tettu tekstimuotoisia paikkamerkkejä, esimerkiksi kuva_tiedosto_1_1, jotka kor-vautuvat ajon aikana dynaamisesti tuotetuilla kuvaajilla ja taulukoilla.

Automaatio toteutetaan Python-kirjastoilla, joista keskeisimmät ovat python-docx ja openpyxl. Ratkaisu poistaa kuvioiden ja taulukoiden manuaalisen kopioi-liitä-vaiheen, yhdenmukaistaa raporttien rakenteen ja ilmeen sekä mahdollistaa do-kumentaation toistettavan tuottamisen osana data-analyysi- ja raportointiputkea. Sisällön ja paikkamerkkien yhteys määrittellään Python-sanakirjarakenteella, jossa kukin paikkamerkki kytkeytyy tiedostopolkuun sekä kuvio- tai taulukkoteks-tiin. Tiedostomappaus muodostuu dynaamisesti kuluvan vuoden perusteella, jo-ten järjestelmä on ajettavissa seuraavina vuosina ilman muutoksia lähdekoodiin. Dokumentin aiemmat versiot jäävät tällöin arkistoon. Kuviossa 32 esitetään pro-cessin kokonaisrakenne lähtöaineistoista valmiin Word-dokumentin tallennuk-seen.



Kuvio 32. Automatisoidun dokumentaatioprosessin kokonaisrakenne.

Dokumentin täyttämisestä vastaa korvauskomponentti, joka käy läpi Word-dokumentin kappalerakenteen, tunnistaa paikkamerkit ja korvaa ne PNG-kuvaajilla tai Excel-taulukoista generoidulla sisällöllä. Korvaukset suoritetaan käänteisessä järjestyksessä, jotta XML-rakenteen indeksit säilyvät eheinä. Kuvaajat lisätään dokumenttiin python-docx-kirjaston `add_picture`-funktiolla. Järjestelmä lukee ensin kuvan koon ja aspektisuhteen PIL-kirjastolla ja skaalaa kuvan enintään 15×20 cm:n kokoiseksi. Kuvateksti sijoittuu suoraan kuvaajan alapuolelle XML-rakennetta muokkaamalla.

Taulukot luetaan Excel-tiedostoista openpyxl-kirjastolla välilehti ja solualue kerrallaan. Numeeriset arvot muotoillaan suomalaisen lokaalin mukaisesti, jolloin desimaalierottimena on pilkku ja tuhaterottimena välilyönti. Taulukko viedään dokumenttiin `add_table`-funktiolla, ja otsikkoteksti lisätään sen yläpuolelle samaan tapaan kuin kuvateksteissä. Valmis dokumentti tallennetaan tiedostoon OmaisuuDENhallintasuunnitelma.docx, ja sisällysluettelon merkitään päivytystarve, jotta otsikkonumerointi ja sivunumerot päivittyvät dokumentin avautuessa.

6 JOHTOPÄÄTELMÄT JA POHDINTA

Opinnäytetyön päätavoitteena oli kehittää koneoppimiseen perustuva malli, joka automatisoi ja tehostaa vesihuoltolaitoksen omaisuudenhallintasuunnitelman laatimista ja päivittämistä. Tavoite saavutettiin kehittämällä modulaarinen, Python-pohjainen ohjelmistoratkaisu, joka tuottaa lakisääteisiä vaatimuksia vastaavat suunnitelman tietosisällöt tilinpäätös- ja investointidatan pohjalta.

Tutkimuksen perusteella mallin tarvitsemat keskeiset tietolähteet koostuvat vesihuoltolaitoksen tilinpäätös- ja kirjanpilotiedoista. Mallin toiminnan kannalta olennaisiksi syötteiksi riittävät liikevaihto, operatiiviset kustannukset sekä taseen loppusumma, joita täydennetään inflaatiota kuvaavalla apumuuttujalla, kuluttajahintaindeksillä. Lisäksi mallissa hyödynnetään investointitaulukkoa, rahoitus- ja korkoparametreja sekä vesihuollon hinnoittelun rakennetietoja. Aineisto ei edellytä salassa pidettävää dataa, mikä tukee mallin yleistettävyyttä.

Omaisuudenhallintamalli toteutettiin kolmikerroksisena arkkitehtuurina, joka koostuu tiedonkeruu- ja integraatiokerroksesta, laskenta- ja mallinnuskerroksesta sekä tulosten visualisoinnista ja automatisoidusta dokumentoinnista. Koneoppimisen keskeisinä menetelminä hyödynnettiin lineaarista regressioanalyysiä ja N-BEATS-neuroverkkomallia, jotka tuottavat aikasarjaennusteet taloudellista päätöksentekoa varten. Neuroverkkomallin suorituskyky varmistettiin walk-forward-validoinnilla, joka vastaa reaali maailman ennustetilannetta. Kokonaisuus osoitti, että toimiva ratkaisu edellyttää syväoppimisen, taloudellisen mallinnuksen ja järjestelmällisen validointiprosessin yhdistämistä.

Mallin tuottamat ennusteet voidaan kytkeä suoraan niitä koskeviin lakisääteisiin vaatimuksiin siten, että ne tukevat valtioneuvoston asetuksen 1173/2025 3 §:ssä edellytettyjen sisältöjen tuottamista. Ennusteiden pohjalta voidaan muodostaa investointi- ja rahoitustarpeiden yhteenveto, arvioida maksujen muutostarpeita sekä toteuttaa riskinarviointi. Koneopetettu omaisuudenhallintamalli integroi nämä osa-alueet yhtenäiseksi prosessiketjuksi, joka päättyy automaattisesti generoituun Word-dokumenttiin.

Case-vesihuoltolaitoksen skenaariomallinnuksessa laitoksen taloudellinen profiili muuttuu 20 vuoden ennustejaksolla rakenteellisesti. Mikäli vesihuoltolaitos suorittaa mallin ehdottamat toimenpiteet, laitoksen tase vahvistuu, omavaraisuus kasvaa ja maksuvalmiusriski pienenee, mutta samalla pääoman tuotto laskee. Kyse ei ole varsinaisesta ongelmasta, vaan muutoksesta, jossa velkavetoinen ja korkea nimellistuottoa tuottava talous kehittyy omavaraisemmaksi ja vakaammaksi. Malli tekee tämän siirtymän näkyväksi. Riskiprofiilin tarkempi tarkastelu syventää kuvaa, sillä kokonaisriskin taso pysyy suunnilleen ennallaan, mutta sen painopiste siirtyy rahoitusrakenteesta ja maksuvalmiudesta kohti kannattavuutta ja operatiivista tehokkuutta. Riski ei siis poistu vaan muuttaa muotoaan, ja päätöksenteon on mukauduttava tähän. Selkein varoitussignaali liittyy omaisuuden ylläpitoon, koska eri tarkastelutavat antavat erilaisen kuvan saneeraustahdin riittävydestä. Tämä osoittaa, ettei yksittäinen mittari riitä arvioimaan infrastruktuurin pitkän aikavälin uusiutumiskykyä, ja havainto on yleistettävissä laajemminkin vesihuoltosektorille.

Opinnäytetyön tulokset alleviivaavat sitä, että vesihuoltolaitoksen taloudellinen kannattavuus on omaisuudenhallinnan perusedellytys. Kun laitoksen liiketoiminta on kannattavaa, se kykenee huolehtimaan vesihuolto-omaisuudestaan ja toteuttamaan suunnitellut saneeraukset niin, ettei omaisuuden arvo heikkene pitkällä aikavälillä. Kun laitoksella on tarve tehdä uusinvestointeja, sen on nostettava taloudellista kannattavuuttaan riittävälle tasolle uusinvestointien rahoittamiseksi. Artefaktin tuottamat liikevaihtotavoitteet ja riskinarvio antavat laitokselle näkymän siihen, millä tasolla kannattavuuden on oltava, jotta nämä edellytykset täyttyvät pitkällä aikavälillä.

Monitasoinen mallinnus tuottaa lisäarvoa nimenomaan kokonaiskuvan kautta. Yksittäiset tunnusluvut voivat antaa harhaanjohtavan kuvan joko liian myönteiseen tai liian kielteiseen suuntaan, mutta kun kannattavuus, vakavaraisuus, maksuvalmius ja omaisuuden ylläpito asetetaan rinnakkain pitkällä aikajänteellä, laitoksen todellinen taloudellinen logiikka avautuu päätöksentekijälle tavalla, joka tukee sekä strategista suunnittelua että kunnan omistajaohjausta.

Opinnäytetyön keskeinen löydös on se, että vesihuoltolain edellyttämän omaisuudenhallintasuunnitelman tietosisällöt voidaan pitkälti tuottaa automaattisesti

olemassa olevasta tilinpäätös- ja investointiaineistosta, mikä tukee hallituksen esityksen HE 40/2025 vp tavoitetta suunnitelman muodon avoimuudesta. Neuroverkkopohjainen ennustaminen osoittautui toimivaksi inflaatio- ja kustannusaikasarjojen mallintamisessa, ja walk-forward-validoinnilla perusteltiin mallin yleistyskyky. Tulokset korostavat, että neuroverkot soveltuvat sekä lyhyen että pitkän aikavälin taloussuunnitteluun. Hinnoittelu- ja riskinarviomodulit kytkevät ennusteet vesihuoltolain kustannusvastaavuuden periaatteeseen sekä laitoksen taloudellisen tilan analyysiin muodostaen kokonaiskuvan laitoksen taloudellisesta kestävydestä.

Tutkimuksen luotettavuutta tukee menetelmävalintojen perusteltavuus ja artefaktin läpinäkyvä kuvaus. Merkittävin rajoite liittyy kuitenkin aineiston laajuuteen. Validointi nojaa pääosin yhden case-vesihuoltolaitoksen aineistoon, mikä riittää suunnittelutieteellisen konstruktio tutkimuksen demonstrointiin osakeyhtiömuotoisessa vesihuoltolaitoksessa, mutta ei mahdollista tilastollisesti yleispätevien johtopäätösten tekemistä eri laitoskokoluokkien, organisaatio- tai toimintamuotojen välillä. Toinen rajoite koskee pitkän aikavälin inflaatioennustetta, jonka luotettavuus heikkenee ennustehorisontin pidentyessä eikä malli pysty tunnistamaan rakenteellisia muutoksia kuten energiakriisien tai geopoliittisten häiriöiden vaikutuksia. Kolmanneksi artefakti on rajauksensa mukaisesti osakokonaisuus: kuntotutkimus-, huolto- ja kunnossapitosuunnitelmat sekä käyttöliittymä ratkaisut on jätetty tutkimuksen ulkopuolelle, joten artefakti sellaisenaan ei kata koko vesihuoltoasetuksen 3 §:n sisältövaatimusta.

Päätavoite kuitenkin saavutettiin. Artefakti automatisoi prosessin lähtödatan tuonnista valmiin Word-dokumentin muodostamiseen. Erityisen hyvin onnistui eri menetelmien yhdistäminen yhtenäiseksi ja toimivaksi kokonaisuudeksi. Inflaatiomalli on rakennettu nelivaiheiseksi putkeksi, joka tukee mallin selkeyttä ja loogista etenemistä. Kassavirtapohjainen liikevaihtotavoite on johdettu selkeäksi laskennalliseksi ratkaisuksi, mikä mahdollistaa tulosten läpinäkyvän tarkastelun.

Opinnäytetyön käytännön merkitys korostuu vesihuoltolaitosten suunnittelun tukemisessa. Kaikkien vesihuoltolaitosten on laadittava omaisuudenhallintasuunnitelma, ja tämä työ tarjoaa konkreettisen toimintamallin siihen, miten suunnitelman

taloudellinen ja investointitietoon perustuva osuus voidaan tuottaa systemaattisesti ja resurssitehokkaasti. Kehitetty artefakti vastaa erityisesti pienten ja keski suurten laitosten haasteisiin, joissa rajalliset henkilöresurssit ja erityisosaamisen puute ovat tehneet suunnittelutyöstä haastavaa.

Tulosten hyödynnettävyys ilmenee siten, että vesihuoltolaitos voi syöttää omat investointitietonsa malliin ja tuottaa suunnitelman taloussuunnittelua koskevan osion sekä riskinarvion automaation tukemana. Mallin tuottamat numeeriset tulokset eivät yksin riitä päätöksenteon perusteeksi, vaan ne edellyttävät asiantuntija-arvioinnin ja paikallisten olosuhteiden tuntemuksen täydennystä. Artefakti vapauttaa asiantuntijatyöaikaan rutiiniluonteisesta laskennasta varsinaiseen analyysiin ja strategisten valintojen perusteluihin. Lisäksi vesihuoltolain 9 a §:n mukainen tiedoksiantovelvollisuus, 18 §:n mukainen vesihuoltomaksujen perusteltavuus sekä 20 a §:n mukainen toimintakertomusraportointi helpottuvat.

Artefaktin empiirinen validointi tulisi laajentaa useamman case-laitoksen aineistolle, jolloin voidaan tutkia, miten malli skaalautuu pienistä vesihuoltolaitoksista suuriin kaupunkilaitoksiin. Koneoppimismenetelmien vertaileva tutkimus jäi rajoituksissa pintapuoliseksi. Jatkotutkimuksessa N-BEATS-arkkitehtuuria voitaisiin verrata systemaattisesti muihin moderneihin aikasarjaennustamisen malleihin kuten Temporal Fusion Transformer -neuroverkkoarkkitehtuuriin.

Artefaktia tulisi laajentaa kattamaan työssä rajatut osa-alueet. Erityisesti kunto- tutkimus-, huolto- ja kunnossapitosuunnitelmien automatisointi yhdistettynä verkostotietojärjestelmän GIS-aineistoon sekä kuntoluokitusdataan tarjoaisi mahdollisuuden teknis-taloudellisen omaisuudenhallinnan kokonaisratkaisun. Tutkimuksessa esitetty ratkaisumalli on osa vesihuoltosektorin digitaalista murrosta, ja sen täysi vaikuttavuus toteutuu, kun se integroidaan laitosten muuhun digitaaliseen infrastruktuuriin, kuten kunnossapito- ja taloushallintojärjestelmiin sekä viranomaisraportoinnin alustoihin.

LÄHTEET

Alegre, H., & Coelho, S. T. (2013). Infrastructure Asset Management of Urban Water Systems. Teoksessa *Water Supply System Analysis – Selected Topics* (s. 49–73). InTech.

Altman, E. I. (2000). *Predicting financial distress of companies: Revisiting the Z-Score and ZETA® models*. NYU Stern Working Paper.

Amadi-Echendu, J. E., Willett, R., Brown, K., Hope, T., Lee, J., Mathew, J., Vickery, J., & Yang, B. S. (2010). What is engineering asset management? Teoksessa J.E. Amadi-Echendu, K. Brown, R. Willett & J. Mathew (toim.), *Definitions, Concepts and Scope of Engineering Asset Management* (s. 35–62). Springer.

American Water Works Association (2018). *Principles of Water Rates, Fees, and Charges* (8. painos; AWWA Manual M1). Denver, CO: American Water Works Association.

Beitner, J. (ei pvm. a). *Models*. PyTorch Forecasting. Haettu 5.1.2026. <https://pytorch-forecasting.readthedocs.io/en/v1.2.0/models.html>

Beitner, J. (ei pvm. b). *NBeats* [API-dokumentaatio]. PyTorch Forecasting. Haettu 5.1.2026. https://pytorch-forecasting.readthedocs.io/en/v1.2.0/api/pytorch_forecasting.models.nbeats.NBeats.html

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: an empirical study on performance estimation methods. *Machine Learning*, 109(11), 1997–2028. <https://doi.org/10.1007/s10994-020-05910-7>

Deuschle, W. J. (2019). *Undergraduate fundamentals of machine learning* [Bachelor's thesis, Harvard College]. Harvard DASH. <https://dash.harvard.edu/URN-3:HUL.INSTREPOS:37364585>

Euroopan keskuspankki. (ei pvm.). *ECB Data Portal* [Aineisto]. Haettu 5.5.2026. <https://data.ecb.europa.eu/>

Falcon, W. & The PyTorch Lightning team. (2019). *PyTorch Lightning* (versio 2.0) [Tietokoneohjelmisto]. <https://doi.org/10.5281/zenodo.3828935>

Fretes, A. V. C. (2025). DuPont System Analysis in Measuring Financial Performance at PT Unilever Tbk during 2019–2023. *International Journal of Business, Management, and Innovation Review*, 2(4), 41–49.

Galabov, M., & Zlateva, P. (2019). Analysis of the natural disaster impact on the company's economy by a modified DuPont model. *International Multidisciplinary Scientific GeoConference SGEM*, 19(5.3), 53–60. <https://doi.org/10.5593/sgem2019/5.3/S21.007>

Hallituksen esitys eduskunnalle laiksi vesihuoltolain muuttamisesta 40/2025. <https://www.finlex.fi/api/media/government-proposal/896209/mainPdf/main.pdf?timestamp=2025-12-10T08%3A04%3A27.198Z>

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society. A, Mathematical, Physical, and Engineering Sciences*, 454(1971), 903–995. <https://doi.org/10.1098/rspa.1998.0193>

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>

Ikaalisten Vesi Oy. (1.4.2026). *Vesihuollon hinnasto 1.7.2026 alkaen*. Ikaalisten Vesi Oy. <https://www.ikaalistenvesi.fi/content/uploads/2026/04/Hinnasto-1-7-2026.pdf>

Irons, R. (2020). *The fundamental principles of finance* (1st ed.). Routledge.

Jekel, C. F., & Venter, G. (2019). *pwlf: A Python library for fitting 1D continuous piecewise linear functions*. University of Florida; Stellenbosch University.

Joshi, A., & Tiwari, H. (2023). An overview of Python libraries for data science. *Journal of Engineering Technology and Applied Physics*, 5(2), 85–90. <https://doi.org/10.33093/jetap.2023.5.2.10>

Julian, D. (2018). *Deep Learning with Pytorch Quick Start Guide: learn to train and deploy neural network models in Python* (1st edition). Packt.

Kaakinen, M., & Ellonen, N. (ei pvm.). *Regressioanalyysi*. Yhteiskuntatieteellinen tietoaarkisto, Menetelmäopetuksen tietovaranto. <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/regressio/analyysi/>

Kirjanpitoasetus 1339/1997. <https://www.finlex.fi/fi/laki/ajatas/1997/19971339>

Kirjanpitolaki 1336/1997. <https://www.finlex.fi/fi/laki/ajantasa/1997/19971336>

Kirjanpitolautakunta. (2007). *Yleisohje suunnitelman mukaisista poistoista 16.10.2007*. Kauppa- ja teollisuusministeriö.

Knüpfer, S., & Puttonen, V. (2018). *Moderni rahoitus* (10., uudistettu painos.). Alma Talent.

Kummaraka, U., & Srisuradetchai, P. (2025). Monte Carlo Dropout Neural Networks for Forecasting Sinusoidal Time Series: Performance Evaluation and Uncertainty Quantification. *Applied Sciences*, 15(8), Article 4363. <https://doi.org/10.3390/app15084363>

Kuulas, A., Renko, T. & Kuivamäki, R. (2020). *Vesihuollon investointitarpeet vuoteen 2040*. Vesilaitosyhdistyksen monistesarja nro 63. Suomen Vesilaitosyhdistys ry.

Laki vesihuoltolain muuttamisesta 1087/2025. <https://www.finlex.fi/fi/lainsaadanto/saaduskokoelma/2025/1087>

Lewinson, E. (2022). *Python for Finance Cookbook*. Second edition. Packt Publishing Ltd.

Niskanen, J., & Niskanen, M. (2013). *Yritysrahoitus* (7. uud. p.). Edita.

Oreshkin, B. N., Carpov, D., Chapados, N. & Bengio, Y. (2020). N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. Teoksessa *Proceedings of the 8th International Conference on Learning Representations (ICLR 2020)*. <https://openreview.net/forum?id=r1ecqn4YwB>

Paavilainen, J. (2019). *Vesihuoltolaitoksen omaisuudenhallinnan käsikirja*. Vesilaitosyhdistyksen monistesarja nro 55. Suomen Vesilaitosyhdistys ry.

Paranko, J. (2012). *Metallituoteteollisuuden kannattavuuskehitys - Pitkän aikavälin näkökulma* [väitöskirja Tampereen yliopisto]. <https://urn.fi/URN:ISBN:978-952-15-2781-4>

Pokiya, K. (2022). *Piecewise linear regression*. Nirma University.

Radivojac, G., Krčmar, A., & Mekinjić, B. (2021). Comparison of Altman's Z-score model and Altman's Z''-score model on the sample of companies whose shares are included in the Republic of Srpska Stock Exchange Index. *Proceedings of the Faculty of Economics in East Sarajevo / Zbornik Radova Ekonomskog Fakulteta u Istočnom Sarajevu*, 21(22), 11–20. <https://doi.org/10.7251/ZREFIS2122011R>

Ryan, O., Haslbeck, J. M. B., & Waldorp, L. J. (2025). Non-Stationarity in Time-Series Analysis: Modeling Stochastic and Deterministic Trends. *Multivariate Behavioral Research*, 60(3), 556–588. <https://doi.org/10.1080/00273171.2024.2436413>

Saleh, H. (2020). *The Deep Learning with PyTorch Workshop: Build deep neural networks and artificial intelligence applications with PyTorch* (1st ed.). Packt Publishing.

Sbrana, A. & Lima de Castro, P. A. (2024). N-BEATS Perceiver: A novel approach for robust cryptocurrency portfolio forecasting. *Computational Economics*, 64(2), 1047–1081. <https://doi.org/10.1007/s10614-023-10470-8>

scikit-learn. (ei pvm.). 1.17. *Neural network models (supervised)*. scikit-learn. Haettu 5.1.2026. https://scikit-learn.org/stable/modules/neural_networks_supervised.html

Sonabend, R. E. B. (2021). *A theoretical and methodological framework for machine learning in survival analysis* [Doctoral dissertation, University College London]. UCL Discovery. https://discovery.ucl.ac.uk/id/eprint/10129352/1/Raphael_Sonabend_PhD_Thesis.pdf

Sumathi, S., Rajappa, S., Kumar, L. A., & Surekha, P. (2022). *Machine learning for decision sciences with case studies in Python* (First edition.). CRC Press. <https://www.taylorfrancis.com/books/9781003258803>

Tilastokeskus. (2021). *Kuluttajahintaindeksi 2021, huhtikuu*. Suomen virallinen tilasto, Hinnat ja kustannukset. <https://www.tilastokeskus.fi>

Tilastokeskus. (ei pvm.). *StatFin: Kuluttajahintaindeksi* [Aineisto]. Haettu 5.5.2026. https://pxdata.stat.fi/PxWeb/pxweb/fi/StatFin/StatFin__khi/

Torres, M. E., Colominas, M. A., Schlotthauer, G., & Flandrin, P. (2011). A complete ensemble empirical mode decomposition with adaptive noise. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Article 5947265. <https://doi.org/10.1109/ICASSP.2011.5947265>

Tuorila, H., & Saastamoinen, A. (2022). *Uudistuva vesihuolto*. Kilpailu- ja kuluttajaviraston tutkimusraportteja 3/2022. Kilpailu- ja kuluttajavirasto. <https://www.kkv.fi/tutkimus-ja-vaikuttaminen/julkaisut/>

Ugarelli, R., & Sægrov, S. (2022). Infrastructure Asset Management: Historic and Future Perspective for Tools, Risk Assessment, and Digitalization for Competence Building. *Water (Basel)*, 14(8), Article 1236. <https://doi.org/10.3390/w14081236>

Valtioneuvoston asetus vesihuollosta 1173/2025. <https://www.finlex.fi/fi/lainsaadanto/2025/1173>

Velten, K., Schmidt, D. M., & Kahlen, K. (2024). *Mathematical Modeling and Simulation: Introduction for Scientists and Engineers* [e-kirja]. (2nd ed.). John Wiley & Sons, Incorporated.

Vesihuoltolaki 119/2001. <https://www.finlex.fi/fi/lainsaadanto/2001/119>

Zhang, Z. (2025). Financial Performance Analysis of U.S. Grocery Retailers Using DuPont Model. *Highlights in Business, Economics and Management (FMIBM 2025)*, 59, 290–296.

LIITTEET

Liite 1. Euribor 12 kk -viitekoron haku ECB Data Portal -palvelusta

```

from datetime import date
import matplotlib.dates as mdates
import matplotlib.pyplot as plt
import pandas as pd
from ecldata import ecldata

# --- Asetukset ---
# EKP Data Portal -sarja-avain: 12 kk Euribor, kuukausikeskiarvo
SERIES_KEY = "FM.M.U2.EUR.RT.MM.EURIBOR1YD.HSTA"
ALKU = "1999-01" # Euron käyttöönotto
LOPPU = date.today().strftime("%Y-%m") # tähän päivään asti

# --- Haku ---
df_raw = ecldata.get_series(SERIES_KEY, start=ALKU, end=LOPPU)

# Etsi aika- ja arvosarakkeet
t_col = next(c for c in df_raw.columns if "TIME" in c.upper())
v_col = next(c for c in df_raw.columns if "VALUE" in c.upper())

df = (pd.DataFrame({
    "pvm": pd.to_datetime(df_raw[t_col], errors="coerce"),
    "korko": pd.to_numeric(df_raw[v_col], errors="coerce")
}).dropna()
.sort_values("pvm")
.reset_index(drop=True))

# --- Kuvaaja ---
fig, ax = plt.subplots(figsize=(10, 5))

ax.plot(df["pvm"], df["korko"], color="blue", linewidth=1.3,
        label="Euribor 12 kk")
ax.fill_between(df["pvm"], df["korko"], alpha=0.15, color="blue")
ax.axhline(0, color="gray", linewidth=0.6)

ax.set_title("Euribor 12 kk", fontweight="bold")
ax.set_xlabel("Vuosi")
ax.set_ylabel("Korko (%)")
ax.grid(True, linestyle="--", linewidth=0.5, color="lightgray")

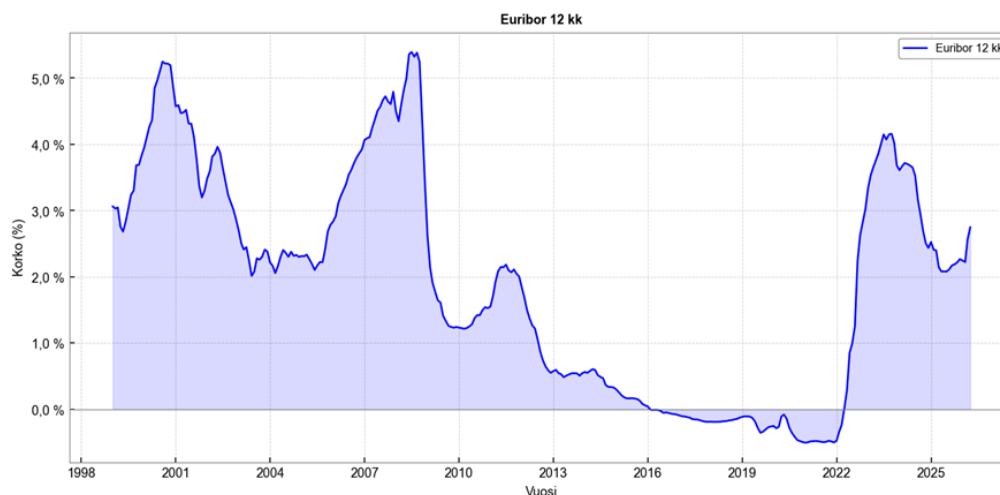
# X-akseli: vuosittain 3 vuoden välein
ax.xaxis.set_major_locator(mdates.YearLocator(3))
ax.xaxis.set_major_formatter(mdates.DateFormatter("%Y"))

# Y-akseli desimaalipilkulla
ax.yaxis.set_major_formatter(
    plt.FuncFormatter(lambda x, _: f"{x:.1f} %".replace(".", ",")))

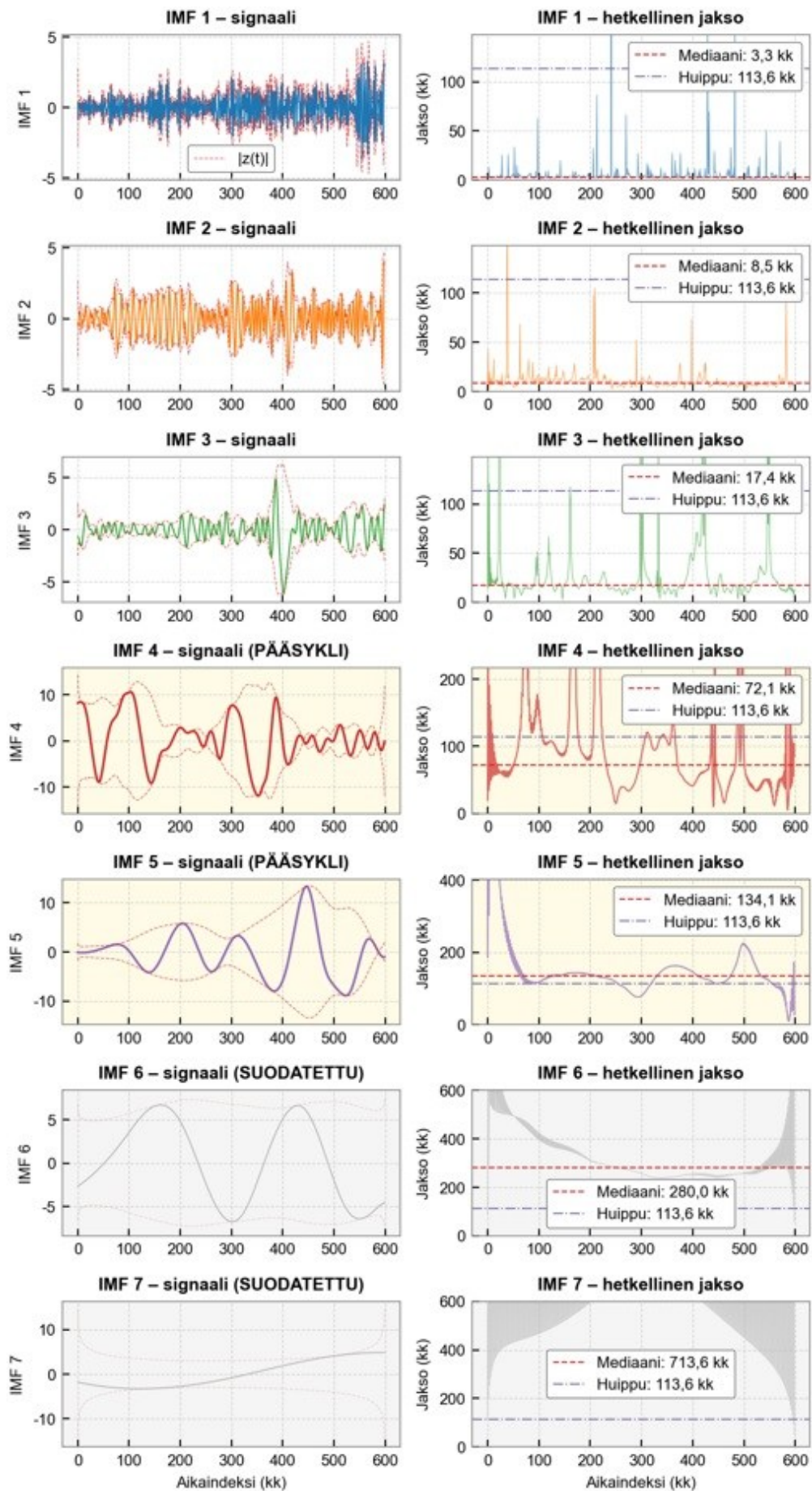
ax.legend(loc="upper right", frameon=True)
plt.tight_layout()

# Tallenna ja näytä
ulos = "euribor_12kk.png"
plt.savefig(ulos, dpi=200, bbox_inches="tight")
plt.show()

```

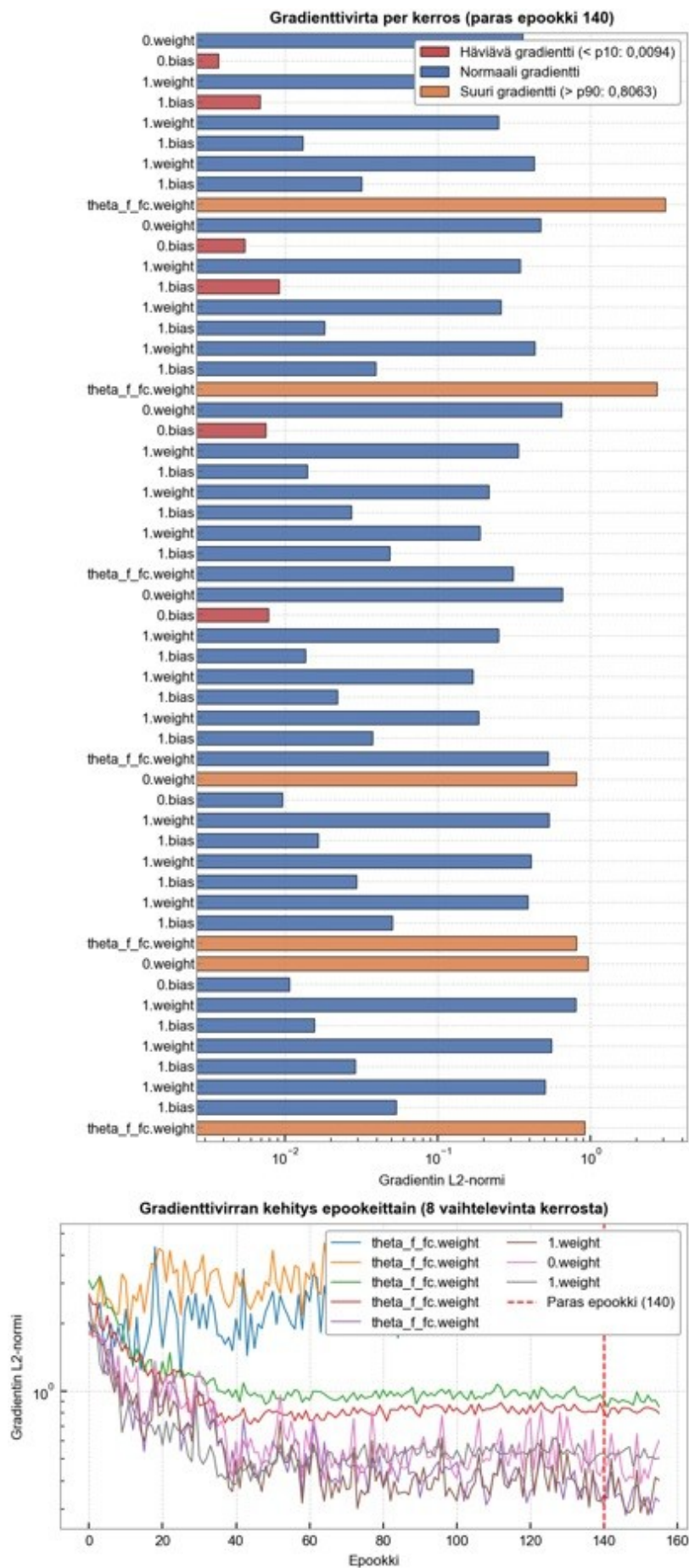


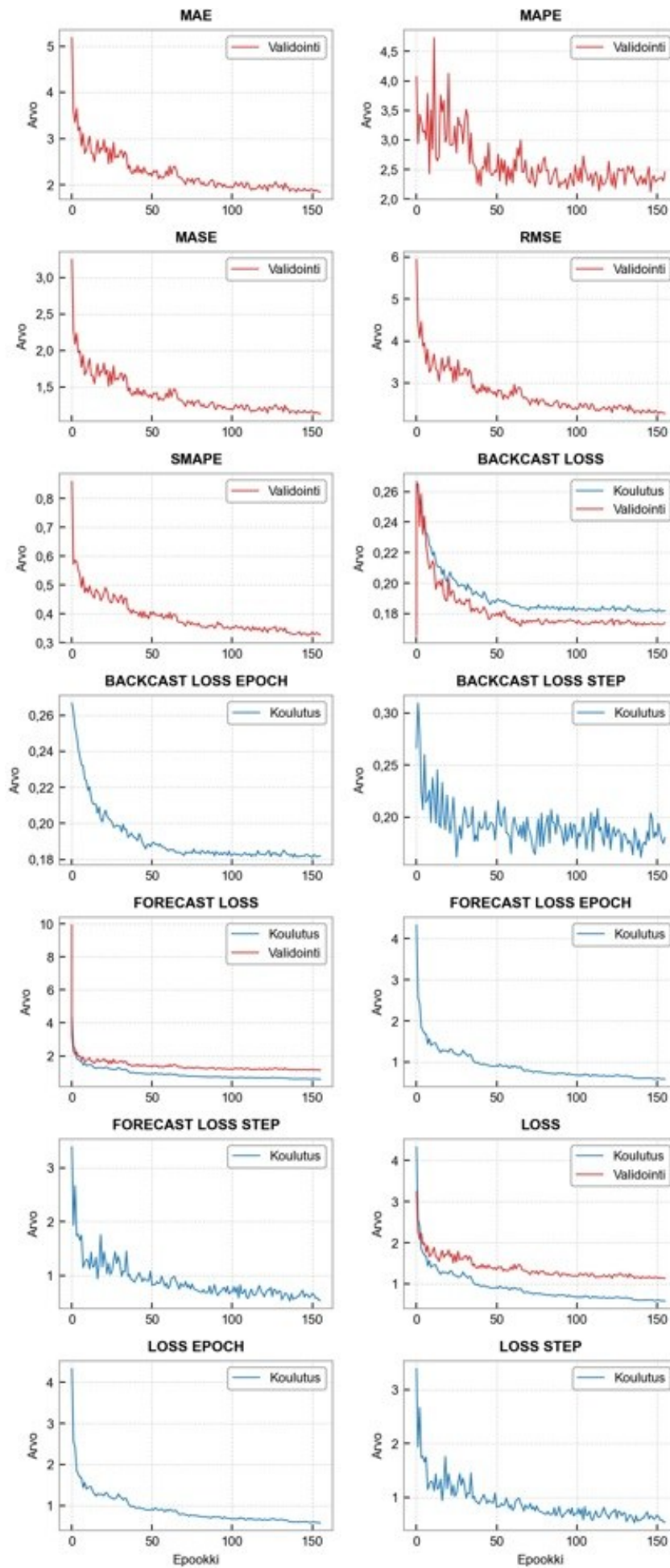
Liite 2. Hilbert IMF -signaalianalyysi



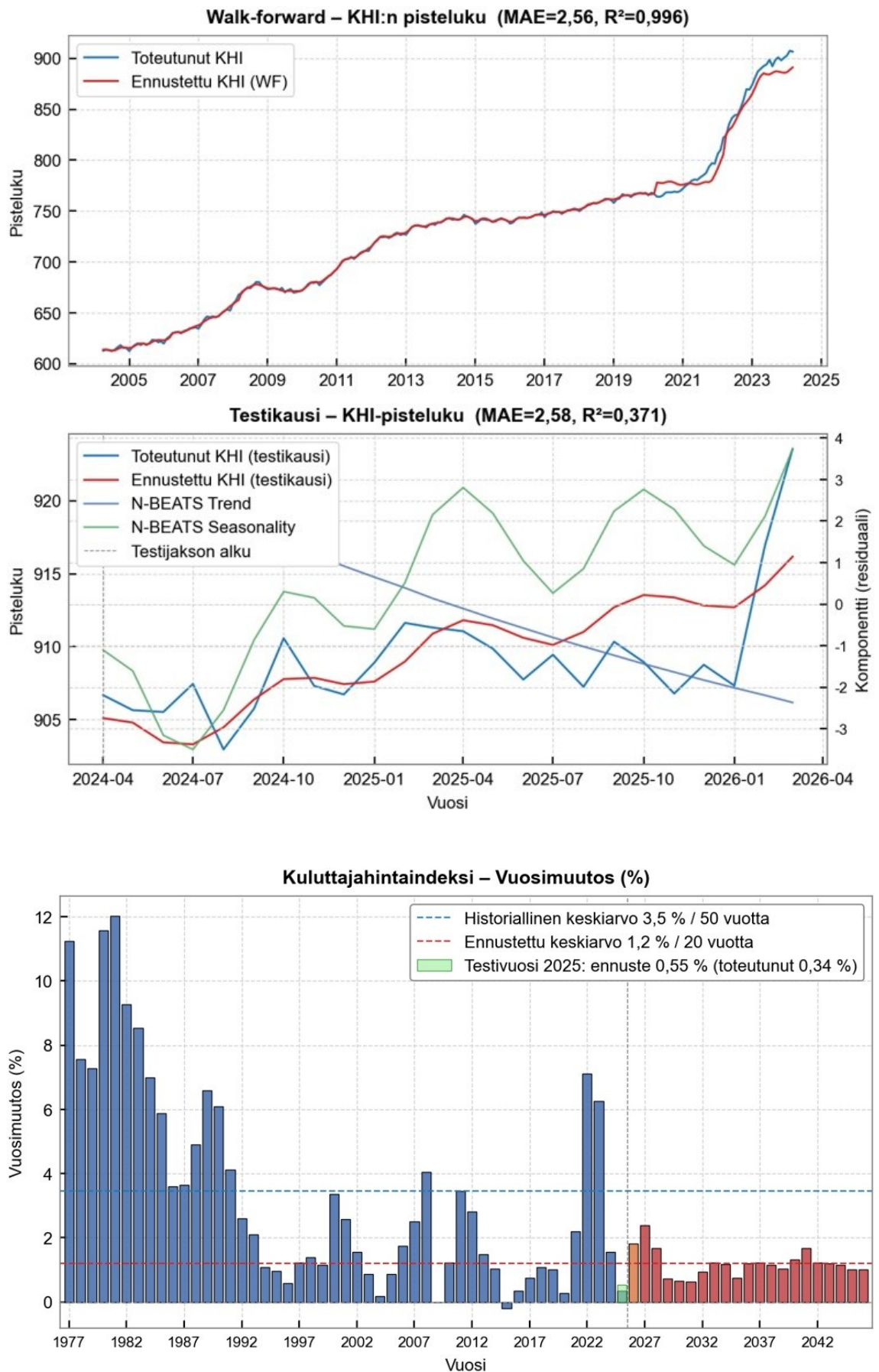
Liite 3. Inflaatiomallin koulutuksen diagnostiikka ja metriikat

1 (2)



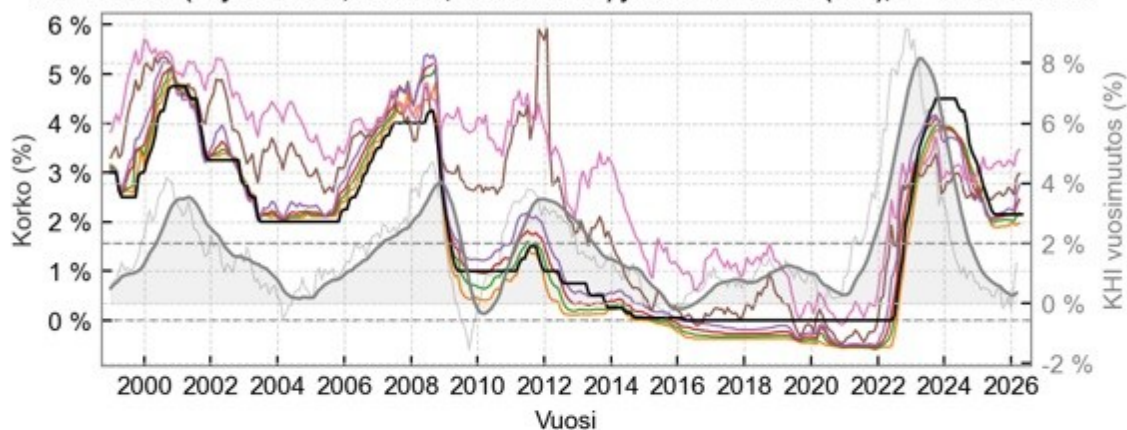


Liite 4. Inflaatiomallin validointijakson ja testikauden ennusteet

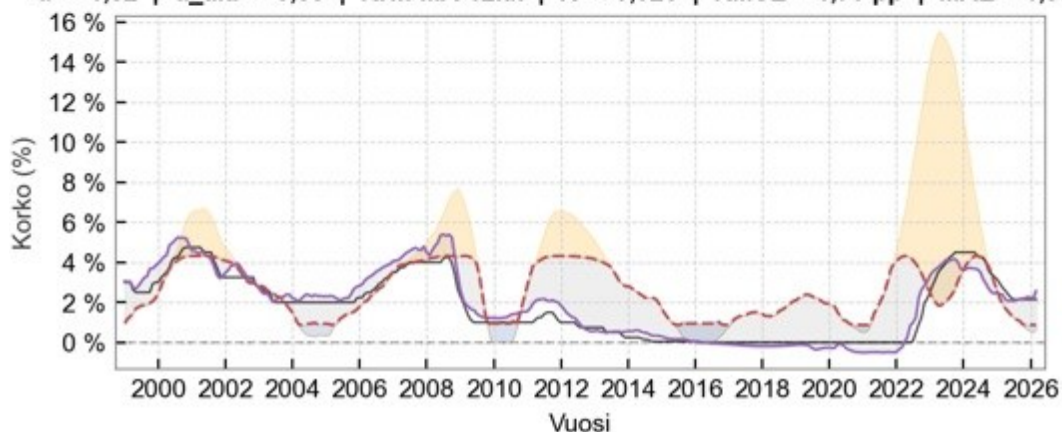


Liite 5. Euribor 12 kk -viitekoron pitkän aikavälin estimointiennuste

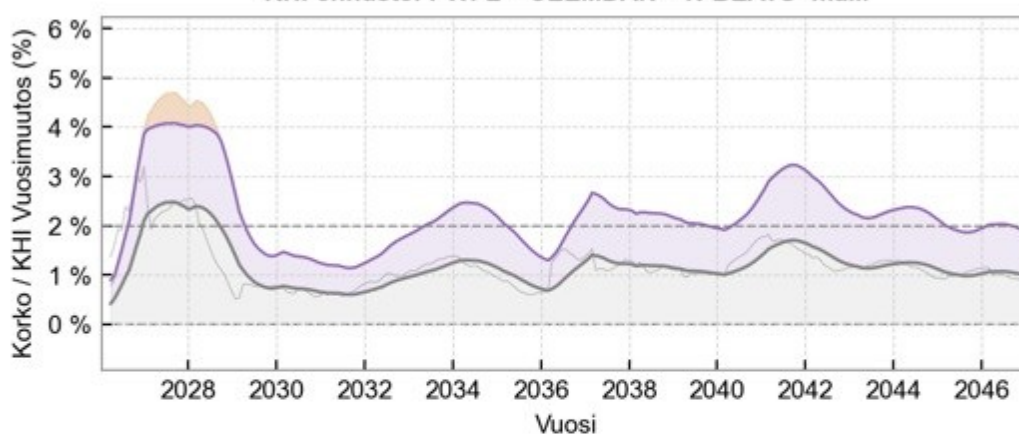
Ohjaus-, Euribor- ja valtionlainakorot ja inflaatio (KHI vuosimuutos-%) 1999–2026
Lähde: EKP (ohjauskorko, Euribor, valtionlainat) ja Tilastokeskus (KHI), 01-1999–05-2026



Mallin validointi historialla: Euribor 12 kk todellinen ja ennustettu (01-1999 - 03-2026)
 $\alpha = -1,32$ | $\alpha_{\text{ala}} = -0,33$ | KHI: MA-12kk | $R^2 = 0,026$ | RMSE = 1,70 pp | MAE = 1,39 pp



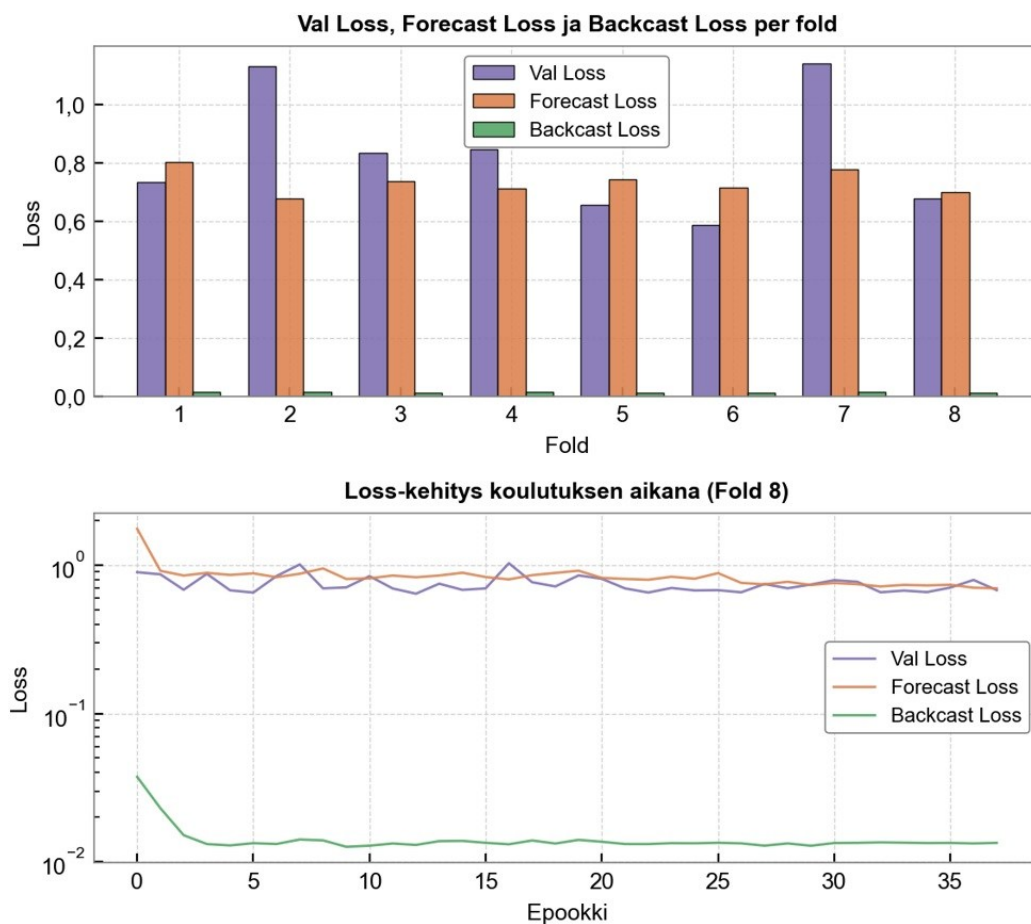
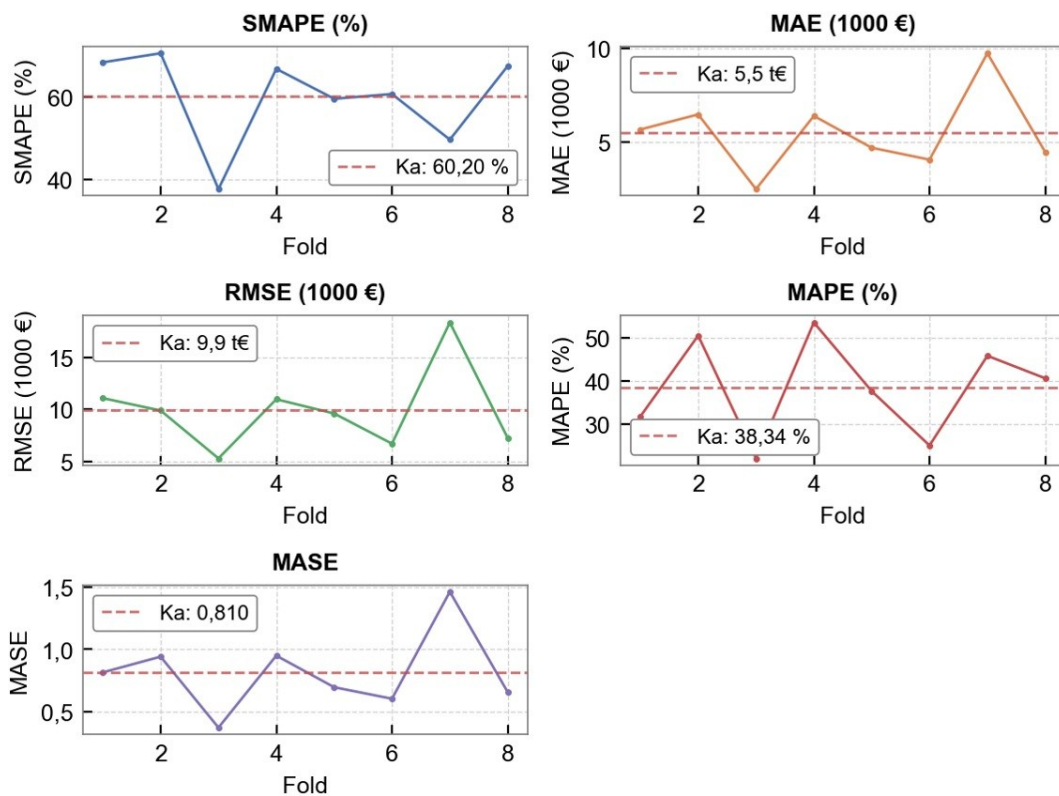
Euribor 12 kk ennuste 2026–2046 (spread=1,8 pp, $\alpha = -1,3$, $\alpha_{\text{ala}} = -0,3$, KHI-tasointus: MA-12kk)
KHI-ennuste: PWFL – CEEMDAN – N-BEATS -malli



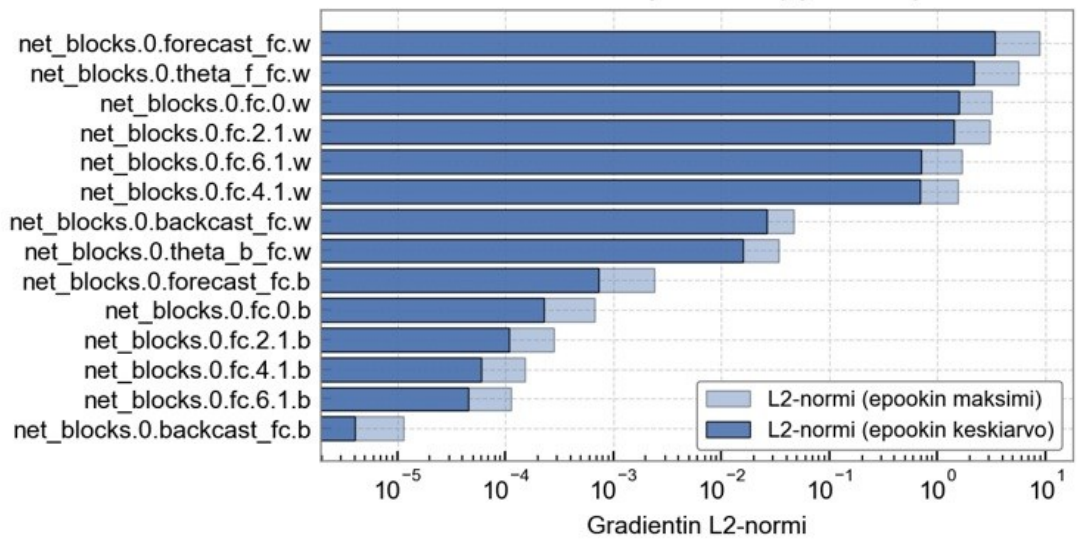
— EKP:n talletuskorko	— Euribor 12 kk, todellinen
— Euribor 1 kk	--- Euribor 12 kk mallin ennuste
— Euribor 3 kk	■ Kiihtyvyysskorjaus yli 2 % (inf.skaalautuva)
— Euribor 6 kk	■ Deflaatiokorjaus alle 0,5 %
— Euribor 12 kk	■ Residuaali (todellinen - ennuste)
— Valtionlaina 5 v	■ KHI:n vuosimuutos (inflaatio)
— Valtionlaina 10 v	■ Spread 1,8 korkopistettä (inf.skaalautuva)
— KHI vuosimuutos (spot)	— KHI:n vuosimuutosennuste (spot)
— KHI vuosimuutos (tasoitettu 12 kk)	— KHI:n vuosimuutosennuste (tasoitettu 12 kk)
--- EKP 2 %:n tavoite	— Euribor 12 kk (ennuste)

Liite 6. Kustannusennustemallin diagnostiikka

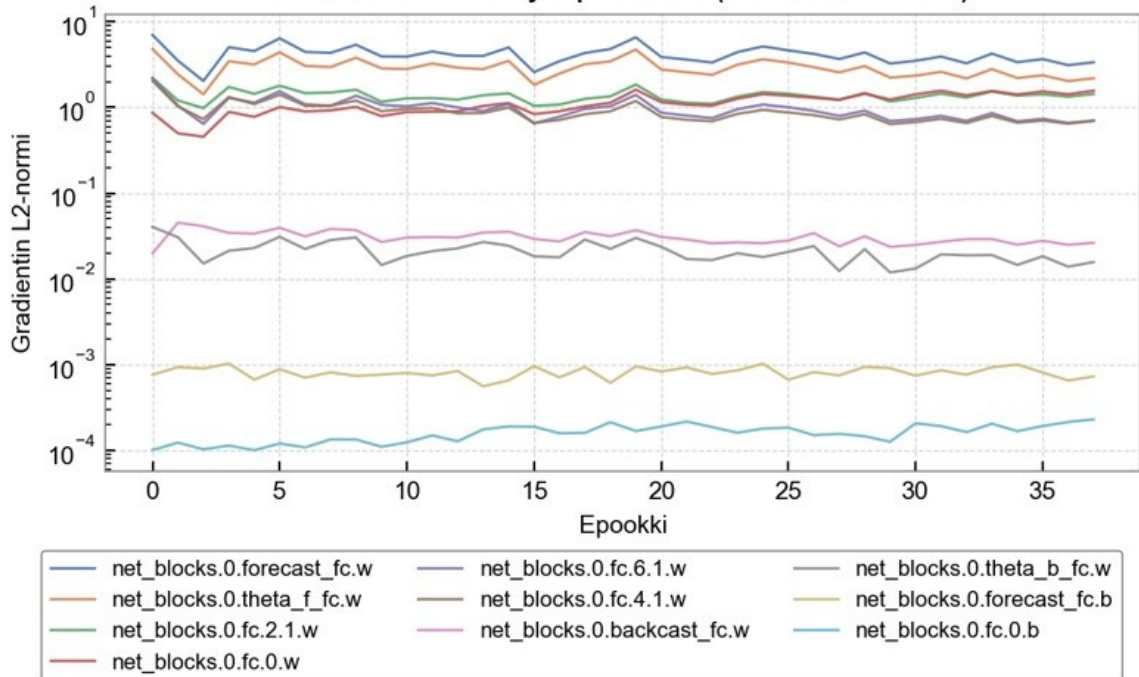
1 (2)



Gradienttivirta per kerros (epookki 37)



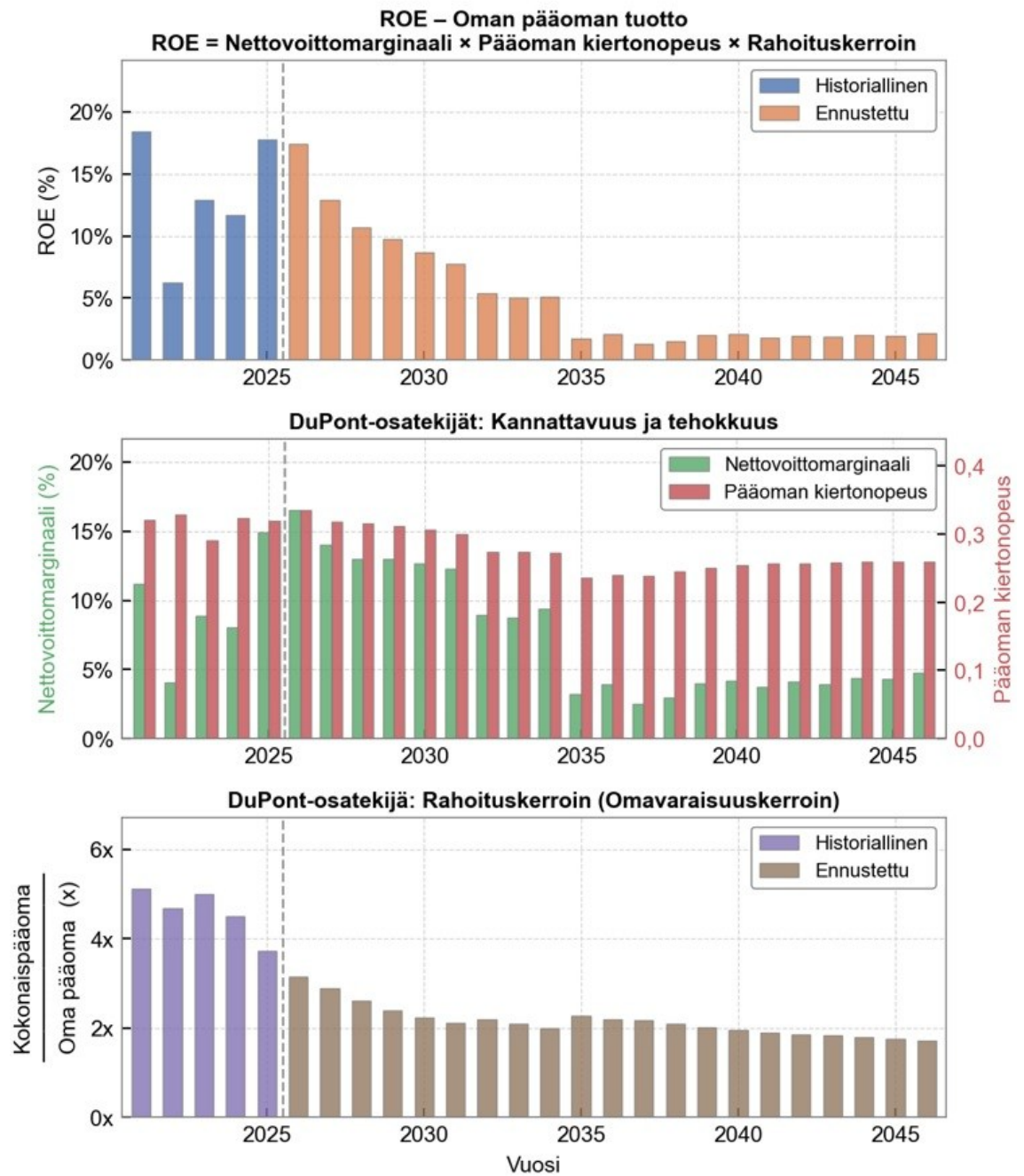
Gradienttivirran kehitys epookeittain (10 suurinta kerrosta)



Liite 7. Riskinarvio – Case-vesihuoltolaitos

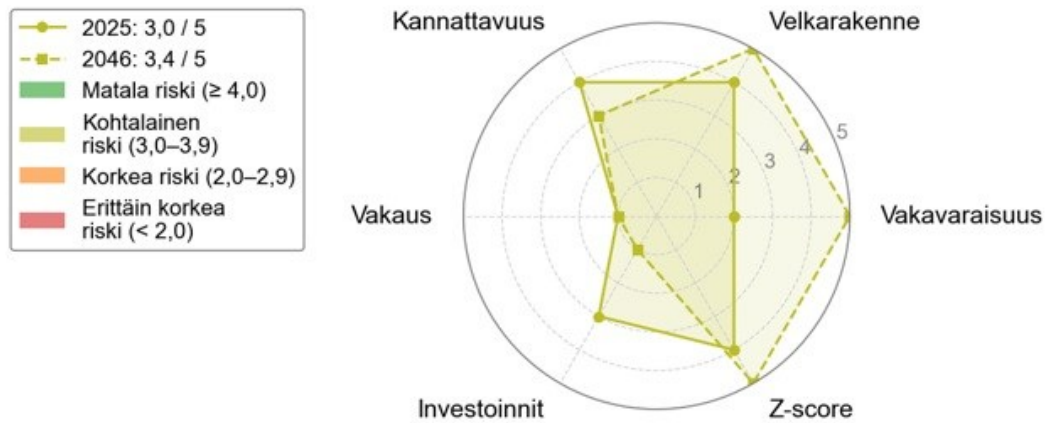
1 (4)

ROE-analyysi – DuPont-kaava

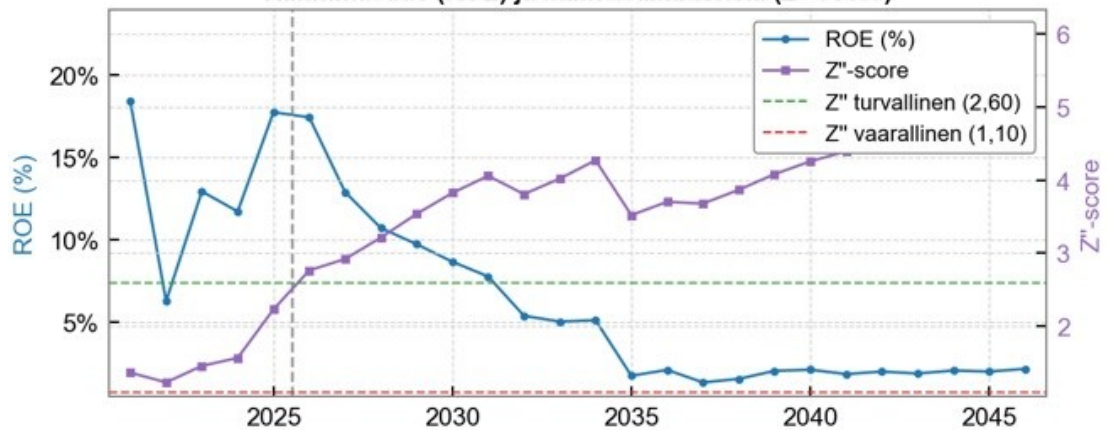


Riskiprofiilianalyysi – Yhteenveto

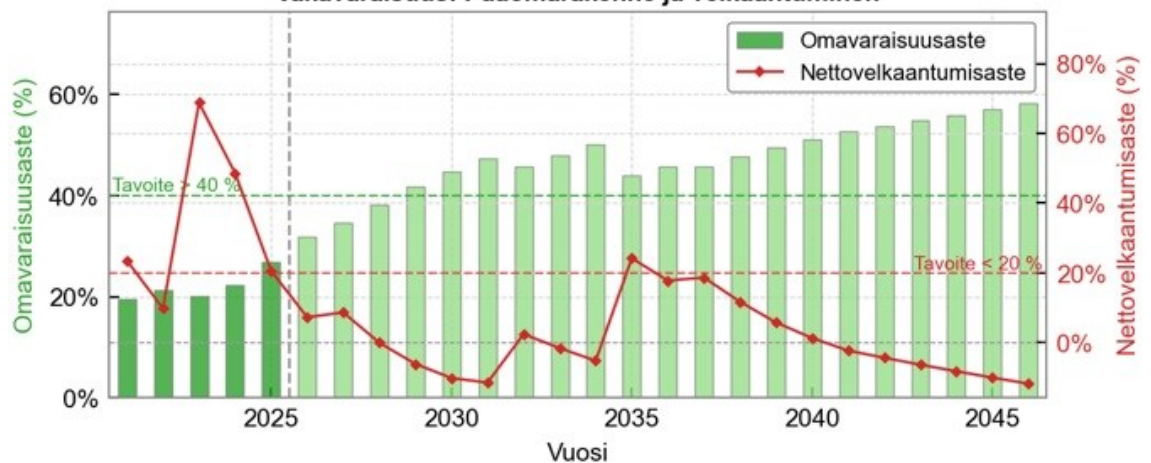
Kokonaisriski: nykyhetki vs. ennustejakson loppu



Kannattavuus (ROE) ja maksuvalmiusriski (Z"-score)

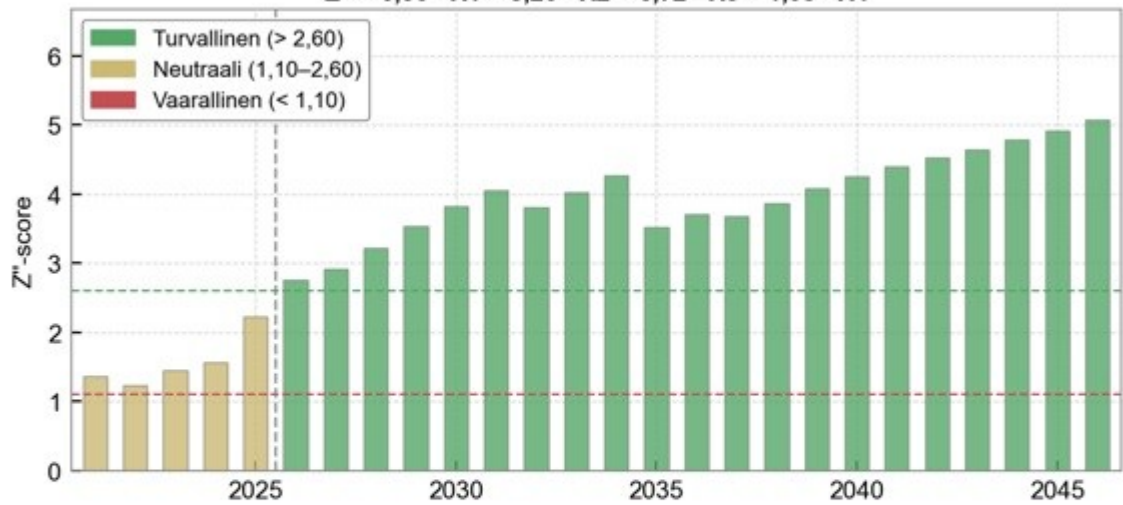


Vakavaraisuus: Pääomarakenne ja velkaantuminen

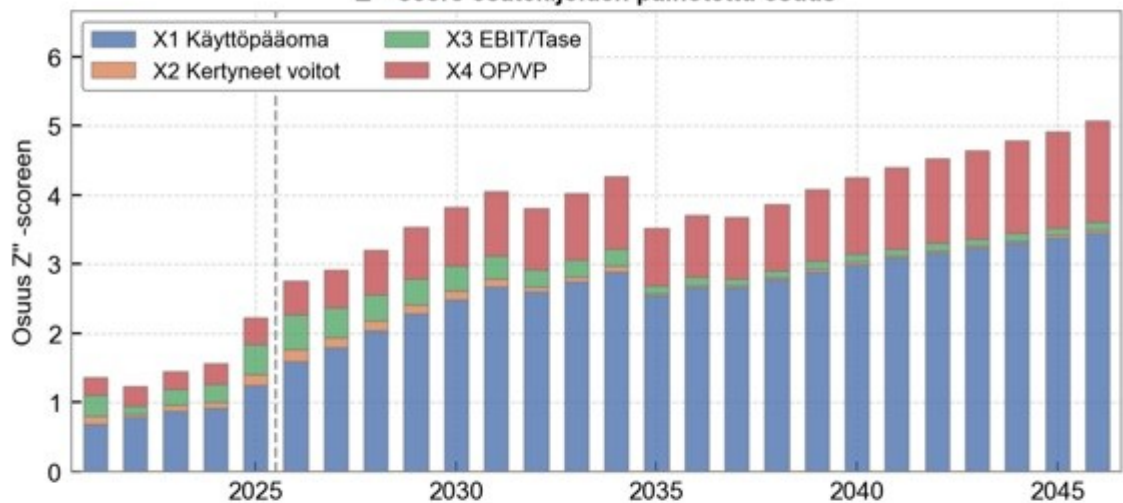


Z"-score ja pääomarakenne

Altman Z"-score vesihuoltolaitos
 $Z'' = 6,56 \cdot X1 + 3,26 \cdot X2 + 6,72 \cdot X3 + 1,05 \cdot X4$

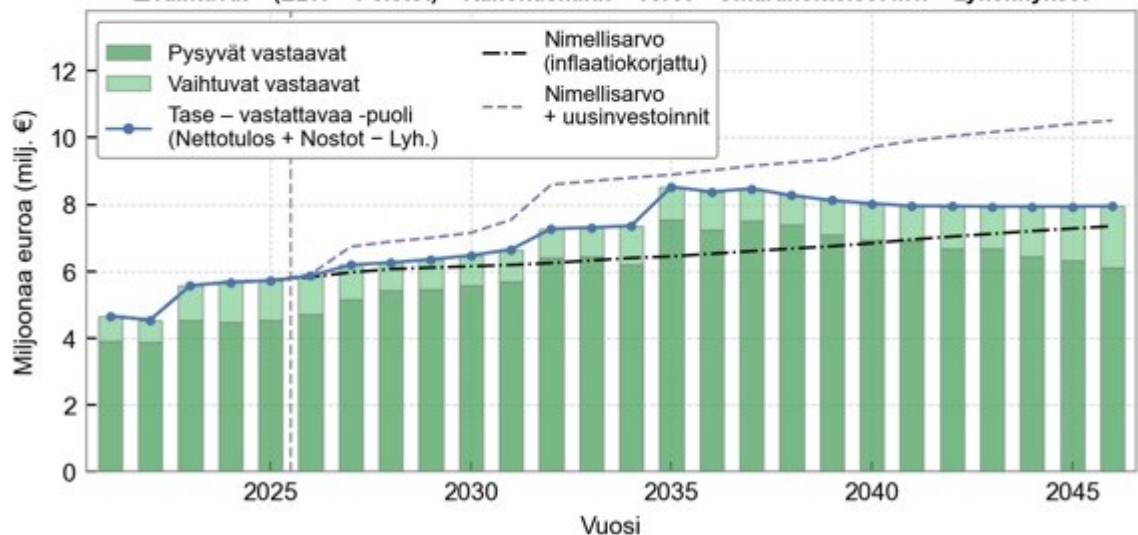


Z"-score osatekijöiden painotettu osuus



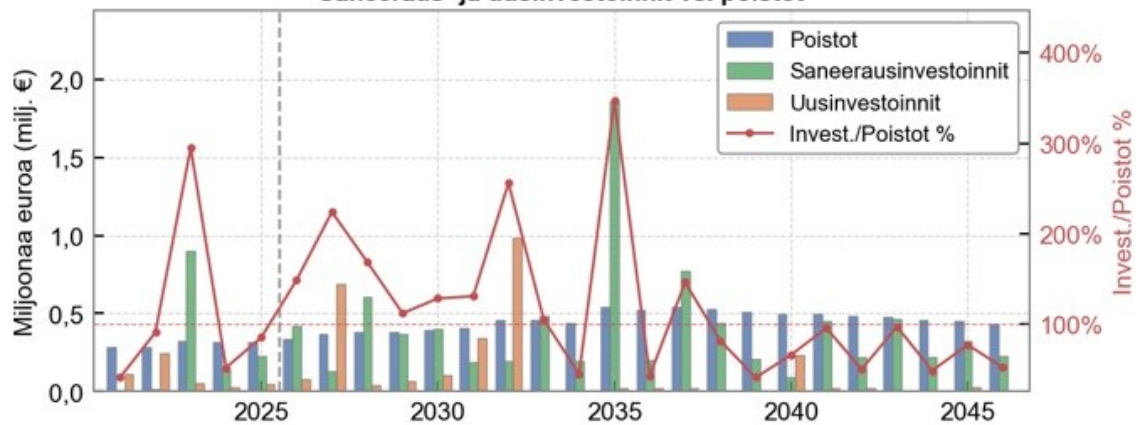
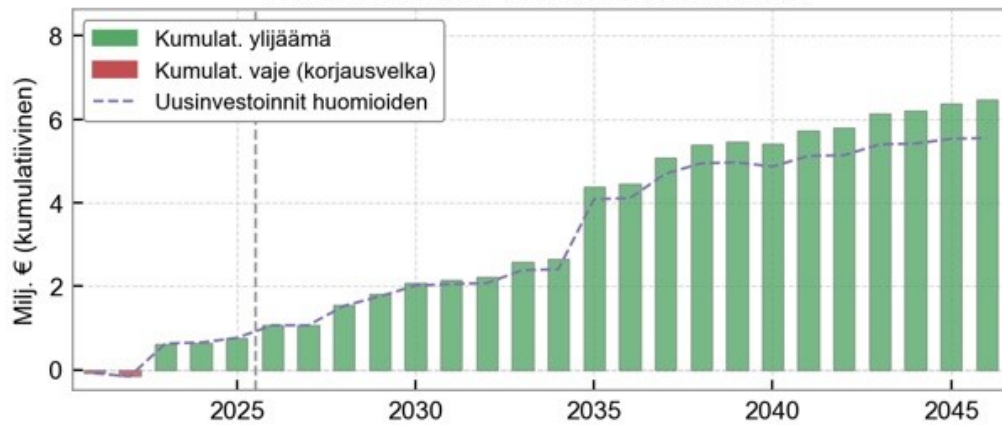
Taseen kasvu (laskennallinen) ja nimellisarvon kehitys

Δ Pysyvät = Investoinnit - Poistot
 Δ Vaihtuvat = (EBIT + Poistot) - Rahoituskulut - Verot - Omarahoitteiset inv. - Lyhennykset



Investointiriski – Omaisuuden ylläpito

Saneeraus- ja uusinvestoinnit vs. poistot

Kumulat. saneerausvaje – inflaatiokorjattu taseen nimellisarvo
Teorettinen tarve = taseen nimellisarvo / 50 vKumulat. saneerausvaje – poistojen taso
Teorettinen tarve = vuosipoistot