Department of Computer Science

Machine Learning-Based Weather Impact Forecasting

Roope Tervo

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 132/2021

Machine Learning-Based Weather Impact Forecasting

Roope Tervo

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held Remote connection link https://aalto.zoom.us/j/69735940472, on 15 November 2021 at 15:00.

Aalto University
School of Science
Department of Computer Science
Machine Learning for Big Data

Supervising professor
Assistant Professor Alexander Jung

Thesis advisor
Assistant Professor Alexander Jung

Preliminary examiners
Associate Professor Tapio Pahikkala, University of Turku, Finland
Researcher Irene Schicker, Zentralanstalt für Meteorologie und Geodynamik, Austria

Opponent
Associate Professor Kai Puolamäki, University of Helsinki, Finland

Aalto University publication series
DOCTORAL DISSERTATIONS 132/2021

ISBN 978-952-64-0524-7 (printed)
ISBN 978-952-64-0525-4 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-64-0525-4

Unigrafia Oy
Helsinki 2021

Finland

NORDIC SWAN ECOLABEL
Printed matter
4041-0619

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Roope Tervo

Name of the doctoral dissertation
Machine Learning-Based Weather Impact Forecasting

Publisher School of Science

Unit Department of Computer Science

Series Aalto University publication series DOCTORAL DISSERTATIONS 132/2021

Field of research Computer Science

Manuscript submitted 28 April 2021 **Date of the defence** 2 November 2021

Permission for public defence granted (date) 15 September 2021 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

Natural disasters influenced over 4 billion people, required 1.23 million lives, and caused almost US$ 3 trillion economic losses between 2000 and 2019. The picture becomes even more deplorable when hazards, smaller-scale severe weather events not requiring casualties, are considered. For example, 78 percent of power outages in Finland were inflicted by extreme weather in 2017, and train delays, often caused by adverse weather, have been estimated to cost 1 billion pounds during 2006 and 2007 in the UK. To mitigate the effects of the adverse weather and increase the resilience of the societies, the World Meteorological Organisation (WMO) raised the consciousness of impact-based warnings along with impact forecasts. Such warnings and predictions can be used in various domains to prepare, alleviate and recuperate from adverse weather conditions.

This thesis studies how to preprocess data and use machine learning to create valuable impact forecasts for power grid and rail traffic operators. The thesis introduces a novel object-oriented method to predict power outages caused by convective storms. The method combines state-of-the-art storm identification, tracking, and nowcasting algorithms with modern machine learning methods. The proposed object-oriented method is also adapted to predict power outages caused by large-scale extratropical storms days ahead. In addition, the thesis studies the task of predicting weather-inflicted train delays. The method presented in the thesis hinges weather parameters on train delays to anticipate the delays days ahead.

The thesis shows that the object-oriented approach is a vindicable method to predict power outages caused by convective storms and that a similar approach is feasible also in the context of extratropical storms. The introduced methods provide power grid operators increasingly accurate outage predictions. The thesis also demonstrates that the train delays related to adverse weather can be predicted with good quality training data. Such predictions offer cardinal information for rail traffic operators in preparing the challenging conditions. Presumably, similar approaches can be applied to any other domain with quantitative impacts produced by identifiable weather events, if sufficient impact data are available. Several advanced machine learning methods were evaluated in the tasks. The results corroborate with existing research: random forests provided a robust performance in all tasks, but also gradient boosting trees, Gaussian processes, and support vector machines proved useful.

Keywords Machine Learning, Weather, Impact Prediction, Supervise Learning

ISBN (printed) 978-952-64-0524-7 **ISBN (pdf)** 978-952-64-0525-4

ISSN (printed) 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2021

Pages 151 **urn** http://urn.fi/URN:ISBN:978-952-64-0525-4

Tekijä
Roope Tervo

Väitöskirjan nimi
Koneoppimiseen perustuvat sään vaikutusennustukset

Julkaisija Perustieteiden korkeakoulu

Yksikkö Tietotekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 132/2021

Tutkimusala Tietotekniikka

Käsikirjoituksen pvm 28.04.2021 **Väitöspäivä** 02.11.2021

Väittelyluvan myöntämispäivä 15.09.2021 **Kieli** Englanti

☐ **Monografia** ☒ **Artikkeliväitöskirja** ☐ **Esseeväitöskirja**

Tiivistelmä

Luonnonkatastrofit vaikuttivat yli 4 miljardiin henkeen, vaativat 1,23 miljoonaa kuolonuhria ja tuottivat lähes 3 biljoonan dollarin taloudelliset tappiot vuosina 2000 -- 2019. Kuva heikkenee entisestään, mikäli huomioidaan myös pienemmän luokan vakavat säätapahtumat. Esimerkiksi 78 prosenttia Suomen vuoden 2017 sähkökatkoista oli sään aiheuttamia. Toisaalta -- usein säähän liittyvät -- junien myöhästymiset tuottivat arviolta miljardin punnan tappiot vuosina 2006 -- 2007 Isossa-Britanniassa. Maailman ilmatieteiden järjestö (WMO) onkin tähdentänyt vaikutusperusteisen varoitusten ja vaikutusennusteiden tärkeyttä vaaralliseen säähän varautumisessa. Vaikutusperusteiset varoitukset ja ennustukset ovat tärkeä apuväline useilla yhteiskunnan osa-alueilla varautuessa ääreviin sääilmiöihin sekä lievittäessä niiden vaikutuksia ja toipuessa niistä.

Tämä väitöskirja tutkii kuinka esiprosessoida dataa ja hyödyntää koneoppmimista sähköverkko- ja junaliikenneoperaattoreille tuotetuissa vaikutusennusteissa. Väitöskirja esittelee uuden oliopohjaisen metodin konvektiivisten rajuilmojen aiheuttamien sähkökatkojen ennustamiseksi. Metodi yhdistää ajantasaiset myrskyn tunnistus-, seuraus- ja lähihetkiennustusalgoritmit moderneihin koneoppimismenetelmiin. Ehdotettu oliopohjainen metodi on myös muokattu ennustamaan laaja-alaisten matalapainemyrskyjen aiheuttamia sähköatkoja. Lisäksi, väitöskirja tutkii sään aiheuttamien junien myöhästymisten ennustamista. Väitöskirjassa esitetty methodi yhdistää sääparametrit junien myöhästymisdataan, jotta myöhästymisiä voidaan ennakoida päiviä etukäteen.

Väitöskirja osoittaa, että oliopohjainen lähestymistapa toimii hyvin konvektiivisten myrskyjen aiheuttamien sähkökatkojen ennustamisessa, ja että vastaavaa metodia voidaan soveltaa myös matalapainemyrskyjen tapauksessa. Väitöskirjassa esitetyt metodit tarjoavat sähköverkko-operaattoreille entistä tarkempia sähkökatkoennusteita. Väitöskirja osoittaa myös, että sään aiheuttamien junien myöhästymisiä voidaan ennustaa mikäli hyvälaatuista koulutusdataa on saatavilla. Tällaiset ennustukset ovat hyvin tärkeitä junaliikenneoperaattoreille haasteellisiin olosuhteisiin varauduttaessa. Oletettavasti samoja lähestymistapoja voidaan hyödyntää myös muilla aloilla, joilla vaikutuksia ovat numeerisesti mallinnettavia ja tunnistettavan säätapahtuman tuottamia sekä kunnollista vaikutusdataa on saatavilla. Väitöskirja vertailee useiden koneoppmismetodeiden soveltuvuutta käsiteltäviin tähtäviin. Tulokset ovat linjassa edellisten tutkimusten kanssa: erityisesti satunnaismetsät ('random forests') tarjosivat toimitavarmoja ennusteita kaikissa tehtävissä, mutta gradienttivahvisteiset puut ('gradient boosting trees'), Gaussiset prosessit ('Gaussian processes') ja tukiverkkokoneet ('support vector machines') toimivat

Avainsanat Koneoppiminen, Vaikutusennustus, Sääennustus, Ohjattu oppiminen

ISBN (painettu) 978-952-64-0524-7 **ISBN (pdf)** 978-952-64-0525-4

ISSN (painettu) 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2021

Sivumäärä 151 **urn** http://urn.fi/URN:ISBN:978-952-64-0525-4

Kiitokset

Pojalleni Aatokselle

Olen aina halunnut tuottaa maailmaan edes pisaran verran uutta tietoa. Toivottavasti tämä toimii esimerkkinä toiveiden toteuttamisesta myös Sinulle.

Haluan kiittää työnantajaani Ilmatieteen laitosta ja pomoani Tarja Riihisaarta tämän mahdollistamisesta. Lisäksi kiitän lämpimästi professoriani Alex Jungia, joka on tarjonnut erinomaista tukea opiskeluuni ja tutkimukseeni. Kanssakirjoittajani Joonas, Ilona ja Laila ovat kontribuoineet merkittävästi tutkimukseen. Kiitän syvästi myös Timoa, Annakaisaa, Anttia, Markoa, Yuta, Laiaa ja Elenaa tekstieni kommentoinnista.

Kiitos äidille, isälle ja isovanhemmilleni, että opetitte arvostamaan sivistystä.

Ja ennen kaikkea, kiitos vaimolleni Tealle tuesta, mitä ilman tämä työ ei olisi ollut mahdollista.

Heidelberg, 15 syyskuuta 2021,

Roope Tervo

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Roope Tervo, Joonas Karjalainen, Alexander Jung. PREDICTING ELECTRICITY OUTAGES CAUSED BY CONVECTIVE STORMS. In *IEEE Data Science Workshop (DSW)*, Lausanne, pp. 145-149, May 2018.

II Roope Tervo, Joonas Karjalainen, Alexander Jung. Short-term prediction of Electricity Outages Caused by Convective Storms. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8618 – 8626, November 2019.

III Roope Tervo, Ilona Láng, Alexander Jung, Antti Mäkelä. Predicting power outages caused by extratropical storms. *Natural Hazards and Earth System Sciences*, 21 pp. 607–627, February 2021.

IV Roope Tervo, Laila Daniels, Alexander Jung. Predicting Weather-Inflicted Rail Traffic Disruptions. Submitted to *Springer Transportation*, 25 pp., April 2021.

Summary of Publications and Author's Contribution

Publication I: "PREDICTING ELECTRICITY OUTAGES CAUSED BY CONVECTIVE STORMS"

Publication I introduces a novel object-oriented power outage prediction method focusing on the outages caused by convective storms. The prediction is based on weather radar data. The method combines storm identification, tracking, and radar-based nowcasting algorithms with state-of-the-art machine learning methods to classify the storm cells based on their damage potential to the power grid. The paper compares random forest classifiers with multilayer perceptron neural networks in the task and finds the random forests to provide better classification results.

The author proposed and designed the method. He also employed background analysis, trained and optimised the machine learning methods, and analysed their performance. The author was the main writer of the paper.

Publication II: "Short-term prediction of Electricity Outages Caused by Convective Storms"

Publication II develops the object-oriented power outage prediction further and deepens the analysis provided in Publication I. The method was also trained with a larger dataset, and the optimisation of the machine learning models was done more assiduously. The paper extends the analysis of the power outages and the performance of the model with an analysis of the importance of predictive features.

The author carried out scientific research and wrote the paper.

Publication III: "Predicting power outages caused by extratropical storms"

Publication III adapts the power outage prediction method, introduced in Publication I and Publication II, to extratropical storms. Similar to the previous publications, the method identifies and tracks the storm objects from gridded weather data and classifies the objects based on their damage potential to the power grid using machine learning. The identification and tracking process is tailored for geographically broad and long-lasting storms. ERA5 reanalysis data were used in the model training. The paper finds Gaussian processes and Support vector classifiers to befit in the classification task.

The author introduced and implemented the method and conducted the scientific analysis. The author was one of two main writers of the paper.

Publication IV: "Predicting Weather-Inflicted Rail Traffic Disruptions"

Publication IV proposes a novel method to predict an average delay of arrived trains at the station based primarily on the weather information. Such method can be deployed to predict the train delays days ahead. The paper introduces binary classification to anticipate severe disruptions along with regression to estimate the amount of delay.

The author designed and developed the method. He also employed background analysis, trained and optimised the machine learning methods, and analysed their performance. The author was the main writer of the paper.

List of Figures

List of Tables

Abbreviations

AMDAR Aircraft Meteorological DAta Relay

ANN Artificial neural network

BART Bayesian additive regression trees

BSS Brier skill score

BT Boosted gradient tree

CAPPI Constant altitude plan position indicator

DBSCAN Density-based spatial clustering of applications with noise

dBZ Radar reflectivity factor

DT Decision tree

ECMWF European Centre for Medium-Range Weather Forecasts

ENS Ensemble decision tree

EPS Ensemble Prediction System

ERA5 ECMWF Reanalysis v5

FMI Finnish meteorological institute

GBC Gradient boosting trees classifier

GBR Gradient boosting trees regression

GBT Gradient boosting trees

GDBSCAN Generalized density-based spatial clustering of applications with noise

GLM Generalised linear models

GNB Gaussian naïve Bayes

GP Gaussian processes

i.i.d. Independent and identically distributed

LR Linear regression

LSTM Long short-term memory network

Luke The Natural Resources Institute Finland

MAE Mean absolute error

MCS Mesoscale convective systems

MLE Maximum likelihood estimation

MLP Multilayer perceptron

NWP Numerical weather prediction

PPI Plan position indicator

QRF Quantile regression forests

RF Random forest

RFC Random forest classifier

RMSE Root mean squared error

RNN Recurrent neural network

SAR Synthetic-aperture radar

SMOTE Synthetic Minority Oversampling Technique

SVC Support vector classifier

TITAN Thunderstorm, Identification, Tracking, Analysis, and Nowcasting

UI User interface

SVM Support vector machine

WMO The World Meteorological Organization

1. Introduction

1.1 Motivation

Adverse weather poses a considerable threat to life, wellbeing, and property and causes tremendous economic loss across the world each year. Between 2000 and 2019, 7 348 major disasters, influencing at least 100 people, requiring more than ten casualties, causing a state emergency, or calling for an international emergency, were reported globally. The disasters influenced altogether over 4 billion people, required 1.23 million lives, and caused almost US$ 3 trillion economic losses. [3]

The picture becomes even more gloomy when hazards, smaller-scale severe weather events, are considered. The comprehensive global statistics have not been gathered, but extensive losses have been estimated in different domains and areas. To name some examples: wind damages forest up to 200 million m^3y^{-1} in Europe, causing considerable losses in economic value, and carbon sequestration [4]; road transport is vulnerable to the extreme weather in UK [5]; 78 percent of power outages in Finland were inflicted by extreme weather in 2017 [6, p.20]; train delays, often caused by adverse weather, have been estimated to cost 1 billion pounds during 2006 and 2007 in UK [7].

Sharp growth in the number of recorded disaster events and the losses have been observed compared to the previous 20 years [3]. Windstorm damages have also increased significantly during the previous decades in Europe [8, 9, 10, 11]. However, the number of windstorms may not be the reason for the increased damages [12, 13, 14] but also the increased vulnerability of society may play a significant role [15].

Many of these consequences could be avoided by anticipating and communicating the events beforehand [2], and especially early warning systems could decrease death tolls and costs [16]. However, this is not always the case, even though the hazardous conditions would have been correctly forecasted. One fundamental reason for the damages is a lack of under-

standing of the impacts of the predicted hydrometeorological conditions among the authorities and the wide audience [1].

In 2015, The World Meteorological Organization (WMO) prioritised the impact-based warnings and impact forecasts for mitigating the consequences induced by the severe weather [1]. Traditionally, meteorological services have provided weather forecasts such as[1]:

> **Forecast for next 24 hours**
> **Sea of Bothnia**
> North to northeast 9-13 m/s, in the evening up to 15. Locally showers of snow, otherwise good vis.

In contrast, the impact-based warnings and forecasts are based on and augmented with the anticipated consequences:[2]

> **Southern Sea of Bothnia**
> **Wind warning for sea areas — valid between 5.2. 09:00 - 5.2. 22:00**
> North to northeast gale 15 m/s starting before noon.
> *Possibly dangerous for recreational crafts and small boats.*

The impact-based warnings typically include the risk level determined from the probability of hazard occurrence with the severity of the hazard with an impact matrix as illustrated in Figure 1.1. If a low impact event occurs with low probability, no warning is given. In a case of high impact with low probability, a low-level warning is given. A high or moderate impact event with high probability triggers a high-level warning. It is notable that the impacts of weather events greatly vary between different areas and societies. For example, a blizzard hardly affects life in northern Lapland but causes chaos in southern European cities.

The impact forecasts and warnings predict the direct consequences. An example of such application is revealed in Figure 1.2. The map shows storm cells categorised based on their damage potential, and the graph tells occurred and forecasted households without electricity. The application is discussed more in Chapter 4. Notably, creating such forecasts require knowledge and data about users' processes, exposure mechanisms, and impact data that can be combined with the meteorological data.

The anticipation of the impacts is not always adequate. Timely and correct mitigating actions need to be taken by various humans and automated agents, such as power grid operators as in Chapter 4 and rail traffic operators as in Chapter 5. Usability and quality of delivered infor-

[1]Example taken from:
`https://en.ilmatieteenlaitos.fi/weather-forecast-for-shipping` at 5/2/2021
[2]Example taken from:
`https://en.ilmatieteenlaitos.fi/warnings` at 5/2/2021

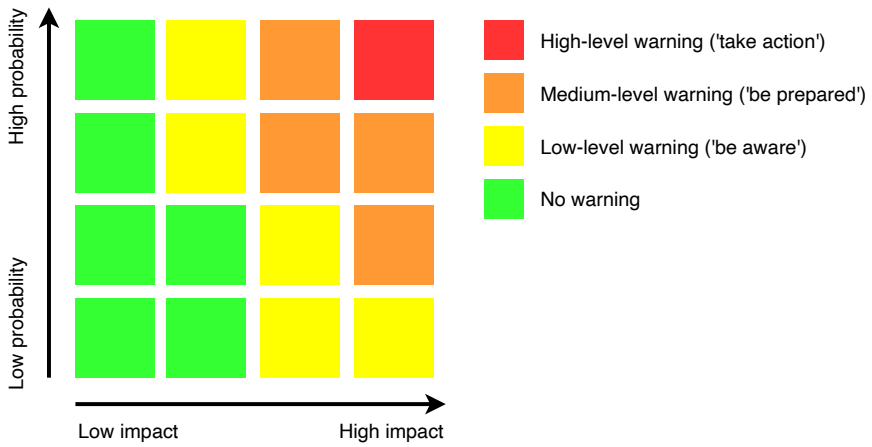

Figure 1.1. Conceptual illustration of the decision process to decide warning level for impact-based warnings, adapted from [1]. The warning level typically depends on the probability of the weather event and the severity of its impacts. This thesis studies methods for computing predictions for the expected impacts of weather.

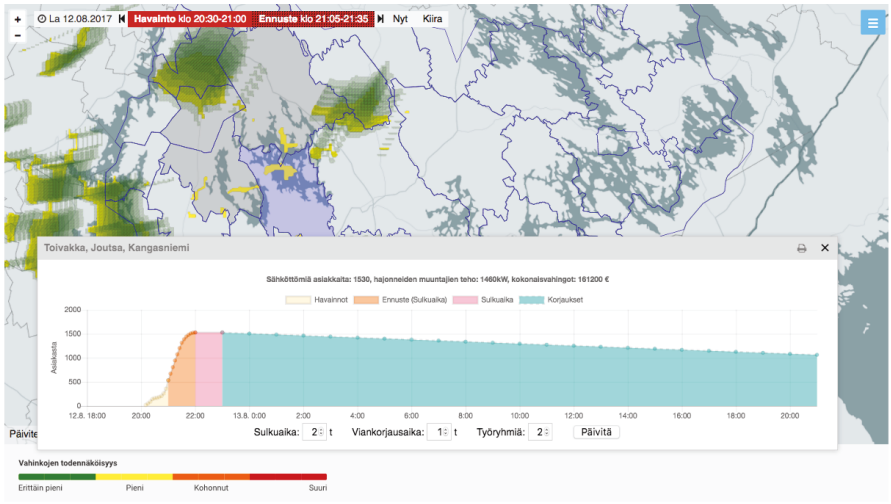

Figure 1.2. An example for an impact-forecast that measures the impacts by the number of households being disconnected from the power grid.

mation depend significantly on the context and requirements of agents receiving the information. Thus, the impact forecasts need to be delivered in different ways and forms for different users [17].

Machine learning is a key technology to create impact forecasts. Generally, machine learning refers to a set of methods capable of learning from data. The methods may be supervised or unsupervised. Supervised methods are trained with correct answers and labels, while unsupervised methods try to learn data structure independently. Impact forecasts typically employ supervised learning with weather data as predictive features and impact data as labels.

Machine learning methods have widely been used, especially for indirect impact predictions, focusing on providing tools for meteorological duty officers and other authorities who can then communicate the warnings to the end-users. Typical domains are for example predicting floods [18], detecting forest fires [19, 20, 21], tropical cyclone modelling [22] and detection [23], classification of convective storms [24, 25], and predict extreme precipitation and hails [26, 27]. Convolutional neural networks have also been applied to find extreme weather from climate datasets [28] and weather datasets [29, 30]. Numerous products providing direct impact forecasts exist as well. Example of such products are predicting flight delays [31], detecting turbulence from weather radar data [32], reducing traffic accidents [33, 34], and predicting power outages [35]. The impacts of weather are not always negative. In the field of renewable energy, predicting solar energy production [36, 37] and wind power production are typical tasks solved with machine learning techniques [38].

As previous examples show, a lot has already been done to predict the impact of adverse weather. However, the existing examples contend with only specific domains in societies. Geographical and infrastructural features also vary, and the methods need to be adapted to dedicated regions and societies. A spectrum of possible machine learning methods is also enormous, and underlying data can be processed and employed in numerous ways. Different methods are often required to create predictions in various time scales. Underlying data and its preprocessing differ in structure, accuracy, and update frequency depending on the time scale. Thus, a dedicated set of methods are required to create a prediction hours ahead, days ahead, or even years ahead.

As an example of a particular domain of impact forecasts, extreme weather events cause a substantial number of power outages each year. While hurricanes are one primary reason for large blackouts [39], especially overhead power lines are vulnerable to heavy winds and lightning caused by convective thunderstorms and extratropical storms in Europe [11]. For example in Finland, hundreds of thousands of households are experiencing power outages caused by weather events each year [6]. These events can not be prevented, and burying all lines underground would

not be economically meaningful in rural areas, but predicting the power outages as accurately and early as possible would help power grid operators to mitigate their impact and provide civil protection authorities the possibility to prepare the society to possible outbreaks. However, the geospatial and temporal resolution and the accuracy of the current power outage predictions need still to be improved for a successful preparedness.

The adverse weather is also a common source of train delays causing significant economic loss each year [7]. Especially intense snowfall and heavy wind inflict delays in northern Europe [40, 41, 42, 43, 44] and China [45]. Although weather conditions have been included in several short-term delay propagation estimations (i.e., [46, 47]), days-ahead predictions of the weather-inflicted train delays have not been made before.

This thesis studies how supervised machine learning can be utilised to create valuable impact forecasts for end-users in northern Europe. Three particular applications are considered: (1) predicting power outages caused by convective storms (Section 4.2.1), and (2) extratropical storms (Section 4.2.2), (3) anticipating train delays caused by adverse weather (Chapter 5). The characteristic geospatial and temporal scale of the events considered in this thesis are illustrated in Figure 1.3. The lead time of the forecasts varies from hours (convective storms) to days (extratropical storms and adverse weather for rail traffic). The event size varies from a kilometre to hundreds of kilometres, correspondingly.

The challenge in creating accurate impact predictions increases from applications (1) to (3). Power outages caused by convective storms are easier to predict than extratropical storms (in the means of machine learning) since the spatial and temporal accuracy of available weather data is better for local events. The train delays are particularly challenging to predict since, in addition to the lower accuracy of the data, they are not typically inflicted by any single identifiable event but long-lasting weather conditions.

1.2 Scientific Contribution

The thesis has three important scientific contributions. First, it proposes a new object-orientated method to predict power outages caused by convective and extratropical storms. The method contains a storm identification with a solid threshold, tracking, feature extraction, and classification based on the damage potential of the storm cells. The proposed method is an unprecedented combination of object-orientated data processing and state-of-the-art machine learning classification methods. Opposite to grid- and point-based methods, the object-oriented methods process the storms as polygons (‘objects’). Most importantly, the combination incorporates merging storm objects with meteorological and non-meteorological features

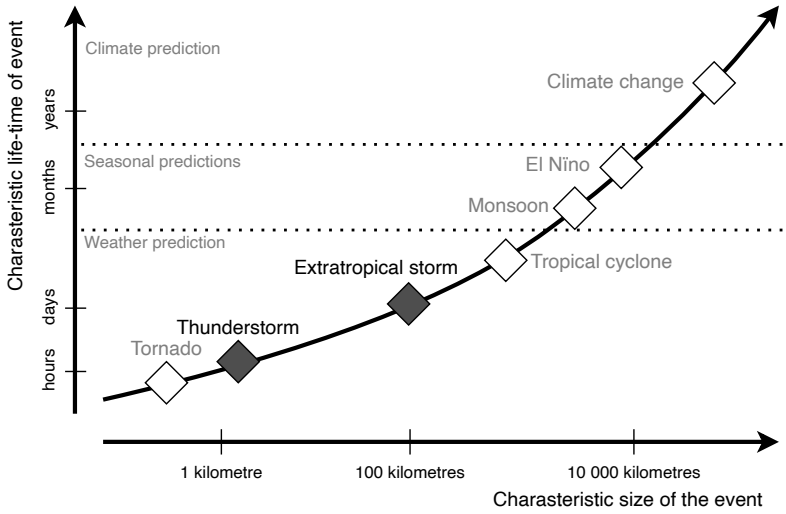

Figure 1.3. Taxonomy of weather events based on their spatial and temporal extend. This thesis considers the events described in dark font. The figure is modified from [2]

and classifying them based on their potential to induce power outages. The method is applied first to the convective storms (Publication I and Publication II) and then modified to the extratropical storms (Publication III).

Second, the thesis introduces a novel framework to anticipate weather-inflicted train delays days ahead (Publication IV). The thesis shows that the weather-related train delays can be predicted days ahead with valuable accuracy. A thorough literature search did not yield previous works providing weather-related long-term train delay predictions.

Third, the thesis delivers essential information on how different machine learning methods function with the weather, power outage, and rail traffic information. It compares several state-of-the-art machine learning methods exploiting the storm objects in the power outage predictions and aggregated ground observations in the train delay prediction.

1.3 Structure

Chapter 2 provides a brief overview of monitoring and predicting adverse weather. Chapter 3 continues with a similar overview on machine learning as relevant for this thesis. Chapter 4 presents methods to forecast weather-related power outages caused by convective storms (Section 4.2.1) and extratropical storms (Section 4.2.2). Chapter 5 presents a method to forecast weather-induced rail traffic disruptions days ahead. Chapter 6 discusses the key results and speculates about future research avenues.

2. Monitoring and Predicting Adverse Weather

While surface weather observations work as the backbone of the global observing system [48], diverse sensors and systems contribute to monitoring the atmosphere. The surface weather observations provide information at a specific location. They are a relevant source of information for applications requiring a time series of faced weather conditions on the Earth surface, such as train delay prediction discussed in Chapter 5.

Some phenomenons, such as storm events, require, nevertheless, more complex instrumentation. This chapter briefly introduces the monitoring and prediction system, essential to this research, for the readers from non-meteorological domains. The particular focus is on convective and extratropical storms causing power outages, discussed more in Chapter 4.

2.1 Convective Storms

Thunderstorms – caused by upward atmospheric motion, *convection* – are typically geospatially small and evanescent phenomena. Individual storms may be only a few kilometres large and endure only tens of minutes [49, 50]. While multi-cell storms or mesoscale convective systems (MCS) may cover large areas, the thunderstorms may also occur as individual, single-cell storms. They often include extremely heavy precipitation, lightning, hails, and heavy winds and can cause severe damage [51, 52].

Being short-living and local events, thunderstorms are hard to observe with sparse surface measurements and model with numerical weather prediction models (NWP). They are most typically monitored with weather radars that scan the atmosphere with a few hundred kilometre range from different elevation angles and produce 3-dimensional volumes of data [53] including several parameters[1] such as radial Doppler speed and radar reflectivity factor (dBZ). Most of the observed radar parameters can also be examined in a two-dimensional plane projection at a certain altitude,

[1] In the context of weather radars, the typical term is *variable*. Here, the term *parameter* is used for the consistence.

called constant altitude plan position indicator (CAPPI) product.

In addition to the regular grid, the convective weather can be analysed in an object-oriented way. The paradigm roots back to the 1970s, when Crane proposed adaptive thresholds to identify and track the storm objects [54]. In the 1990s, Dixon and Wiener introduced a widely used *Thunderstorm, Identification, Tracking, Analysis, and Nowcasting* (TITAN) algorithm, which identifies the objects from 3-dimensional data [55]. Several studies have further extended and developed the object-oriented storm identification and tracking (i.e. [56, 57, 58, 59, 60, 61]). All proposed algorithms identify the storm cells from the regular grid by finding 2- or 3-dimensional contours with a single [58, 60, 55], adaptive [54] or multiple thresholds [57, 56]. The tracking of the identified storm cells may be conducted as the combinatorial optimisation problem [62] like in TITAN [55] or based on the heuristic criteria such as geospatial proximity or overlapping [56, 57, 58, 60, 61]. The object-oriented tracking can be further improved by moving the previous time step objects with estimated velocity vectors before applying the heuristic criteria. A split and merge of the convective storms may be considered as well (i.e., [57, 60]).

A few-hours-ahead prediction of the precipitation and convective storms is typically conducted with a *nowcast*. Nowcasts are often, but not always, based on the weather radar data, and radar-based nowcasts focus on the precipitation. Traditionally, they are based on the velocity vector field, calculated to the whole grid employing cross-correlation (i.e., [63]) or optical flow algorithm [64], which is used to extrapolate the observed precipitation pattern forward in time (i.e., [65]). The methods have been further improved to take a stochastic nature of the precipitation into account and employ NWP data (i.e., [66]). The object-oriented methods provide a computationally less consuming way to predict the storm motion since instead of defining the velocity vectors for the whole grid, the vectors and the extrapolation are determined only for the identified storm cells. For example, TITAN [55] identifies the storm cells from weather radar data, tracks them, and provides a short-term prediction of their movement. Recently, the artificial neural networks are widely experimented in the nowcasting (i.e. [67, 68, 69]). For a more detailed review, the reader is asked to consult [70].

Figure 2.1 instantiates the data types one needs to process when dealing with the radar and nowcasting data. Radar data are typically processed in a regular grid and further handled as images. Thus, radar data cover the whole spatial domain, while ground observations are available only at specific points. Notably, instrumentation varies at each observation station, and most stations observe only a subset of all observable parameters.

The method presented in Section 4.2.1 exploits the operative CAPPI radar data produced by the Finnish Meteorological Institute (FMI). The data is cleaned from non-meteorological echoes with an anomaly removal,

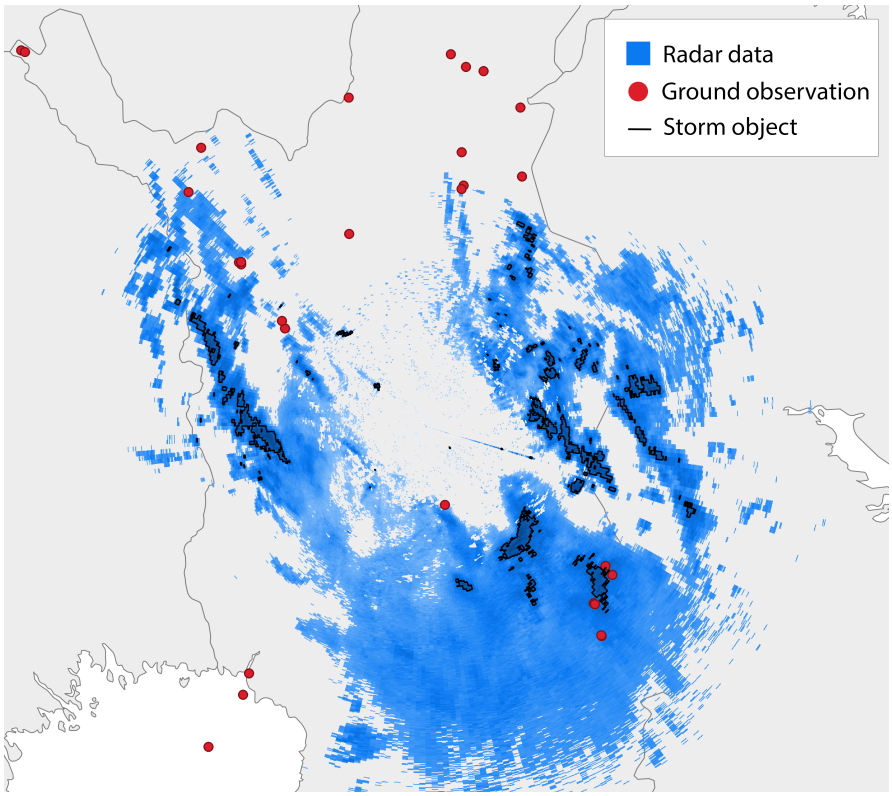

Figure 2.1. Illustrative example of data types in the thunderstorms-inflicted power outage predictions in Section 4.2.1. The blue areas in the figure represent the precipitation. Identified storm objects, solid polygons with intensive precipitation, are marked with a black line.

described in [71]. The weather radar parameter used in this work is the radar reflectivity factor, a quantity based on the backscattered power from the atmospheric rain or snow particles, which can be converted to represent the rainfall intensity. The data are derived from 11 Doppler C-band radars covering Finland, which scan in a spatial resolution of 250 metres and a temporal resolution of 5 minutes.

2.2 Extratropical Storms

Extratropical storms, also referred to as extratropical cyclones, are caused by synoptic-scale low-pressure areas. They are typically over 1000 kilometres wide, and their lifetime varies from several days to 2 weeks. Extratropical storms appear in the middle latitudes of the Earth and often cause a rapid temperature change, yielding weather fronts from the center of the storms. Such storms can produce heavy winds and blizzards but can also be embedded with thunderstorms and tornadoes." [72]

The process of monitoring and predicting extratropical storms consists of observations, analysis, and prediction phases. In the first phase, ground observations, soundings, aircraft observations (AMDARS), and satellite observations compose an earth observation data feed. In the second phase, the feed is fed to *the assimilation* process, creating *an analysis* representing a state of the atmosphere. The assimilation process typically contains a short-range forecast from the previous analysis time step and the observations correcting the forecast. The resulting new analysis is then used to update the forecast, and the process is repeated. The analysis phase can also be calculated retrospectively to produce a consistent long time series about the past climate in the process called reanalysis. A more detailed description of the assimilation process is available in [73].

The third phase contains creating a weather forecast with a long-term simulation based on the analysis. The typical forecast range is from 2 to 14 days ahead of the analysis time. The second and the third phase together are called *a numerical weather prediction model (NWP)*. The modern weather forecasting systems contain also an ensemble prediction system (EPS), where the long-term simulation is repeated several times with slightly different analyses. The process yields an ensemble of forecasts providing different weather scenarios and can thus provide probabilistic information to the forecast. For more details on a the weather forecasting process, the reader is asked to consult [74].

The analysis and NWP data are stored and processed on the regular grid as illustrated in Figure 2.2. The data is typically produced in 3 dimensions but may be reduced into 2 dimensions on a level of interest in the specific application. A one particular level is the land surface where the parameters are provided from the specific altitudes near the Earth

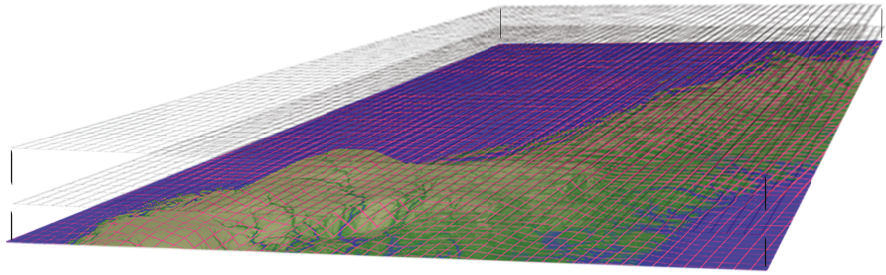

Figure 2.2. NWPs discretise the atmosphere using a regular grid and compute different parameters for each grid point. The dark grids represent the model levels and the red grid the surface level. Instead of using these grid points directly, the methods presented in Chapter 4 extract weather events (storm objects) from the grid.

ground. For example, the temperature is typically provided at a 2 metre altitude, wind at a 10 metre altitude, and snow depth from the ground level.

The method presented in Section 4.2.2 exploits ECMWF ERA5[2] reanalysis land surface data [75] in a model training. The data are in 9 kilometre spatial and 1-hour temporal resolution. Using the model in operations would, however, require changing from reanalysis to NWP data.

[2]ERA5 data may be downloaded from Copernicus Climate Data Store: `https://doi.org/10.24381/cds.bd0915c6`

3. Machine Learning Primer

Machine learning is typically seen as a family of *models* creating predictions[1] based on given *data*. The model consists of the underlying machine learning method, which is fitted to the data based on *a loss function* in the training process. This chapter illuminates the key components and terms of machine learning. The aim is to be concise but comprehensive enough so that the readers not familiar with machine learning can follow the thesis. Machine learning can be *supervised*, *unsupervised*, or *reinforced*. This thesis focuses on supervised learning. For more details, the reader is asked to consult, for example, [76].

3.1 The Concept of Data

Data consist of predictive *features* and, in supervised learning, corresponding *labels*. Each datapoint has n features stacked to a feature vector $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$. Data contain m datapoints composing a matrix $\mathbf{X} = (\mathbf{x_1}, ..., \mathbf{x_m})^T \in \mathbb{R}^{m \times n}$. In the supervised learning, training data are also complemented with m labels, correct answers of the prediction task, composed to a vector $\mathbf{y} = (y_1, ...y_m) \in \mathbb{R}^m$. The primary features $\mathbf{X}$ in this thesis are extracted from weather observations, weather radar data and reanalysis, introduced in Chapter 2. The corresponding labels are discussed in Chapter 4 and 5. Data are often assumed to be *independent and identically distributed* (i.i.d.), which does not hold for temporally and geospatially dependent weather variables. Nevertheless, several studies indicate good results obtained with machine learning models applied to the weather data (i.e., [24, 77, 78]) despite of its internal structure.

[1] It is important to distinguish the term prediction from its ordinary meaning. In the context of machine learning, prediction refers to the outcome of the machine learning model, which is not necessarily anticipated future, such as weather prediction.

3.2 The Concept of Model

The objective of a machine learning task is to find a mapping from features to labels for all possible datapoints:

$$y \approx h(\mathbf{x}) = \hat{y},$$

where y denotes the label, $\hat{y}$ the model prediction, and $h(\mathbf{x})$ refers to the hypothesis. If the task is supervised, y is known for all datapoints used in the training. In unsupervised cases, the correct answers are undetermined. The machine learning task may be *classification* or *regression* depending on whether y is discrete (classification) or not (regression).

The function $h(\mathbf{x})$ is also called *hypothesis*. In practice, the hypothesis is confined to a limited hypothesis space called *model*. The models include always at least one variable vector optimised in the training process. For example, a unregularised linear regressions (LR) is defined as:

$$h(\mathbf{x}) = \hat{\mathbf{y}} = \mathbf{x}^T \mathbf{w}, \tag{3.1}$$

where $\mathbf{w}$ is a weight vector acting as a variable. The training task is then to find optimal values for $\mathbf{w}$ so that $\hat{\mathbf{y}}$ is as close to $\mathbf{y}$ as possible.

The main challenge in the machine learning task is to find an appropriate model that is powerful enough for the particular mapping but is not too powerful. Overly complex models need to be avoided for two reasons: First, as the models gain complexity, they also become computationally unduly expensive to train. Second, when hypothesis space enlarges due to the complex models, the models tend to *overfit*. In other words, they find a perfect solution for the training data but *generalise* poorly to the unprecedented samples. The second problem is also coped with *regularisation*.

3.3 The Loss Function

The model is fitted to the data by minimising *a loss function*. In the supervised learning, the loss function estimates the difference between the prediction $\hat{\mathbf{y}}$ and the label $\mathbf{y}$. Loss functions and their derivatives against weights $\mathbf{w}$ are typically desired to be continuous for an efficient training process. A squared error works as an example of loss function used in the regression:

$$\mathscr{L}\big((\mathbf{x}, y), h\big) := \big(y - h(\mathbf{x})\big)^2 = (y - \hat{y})^2$$

In the classification loss functions, $h(\mathbf{x})$ is often assumed to provide a probability that the sample belongs to the class y. One prevalent example of loss functions is *cross entropy loss*. In the binary classification, the loss is:

$$\mathscr{L}\big((\mathbf{x}, y), h\big) := -\big(y \log\big(h(\mathbf{x})\big) + (1 - y) \log\big(1 - h(\mathbf{x})\big)\big)$$

If multiple classes are predicted, the loss gets the form:

$$\mathscr{L}\big((\mathbf{x}, y), h\big) := -\sum_{c=1}^{C} y_{ind} \log\big(h(\mathbf{x})\big), \tag{3.2}$$

where $y_{ind} = 1$, if $y = \hat{y}$ and 0 else.

As above examples denote, the loss is applied to a simple datapoint. Using the i.i.d. assumption, the expected value of the loss can be approximated with an average loss of data $(\mathbf{X}, \mathbf{y}) \in \mathbb{R}^{m \times n}$. Hence, *an empricial risk*, often called also as *a cost*, is calculated based on the loss:

$$\mathscr{E}\big(h|(\mathbf{X}, \mathbf{y})\big) = \frac{1}{m} \sum_{i=1}^{m} \mathscr{L}\big((\mathbf{x}^{(i)}, y^{(i)}), h\big) = J(\mathbf{w}) \tag{3.3}$$

3.4 Training Process

Training the model can be accomplished using several techniques. One of the prevailing methods is gradient descent, in which the weights are updated gradually so that the cost function is minimised. On each step, the weights get a new value following:

$$\mathbf{w}_{new} = \mathbf{w}_{old} - \alpha \frac{\partial J(\mathbf{w})}{\mathbf{w}}, \tag{3.4}$$

where α is a learning rate and $J(\mathbf{w})$ is a cost function.

A dataset is typically used for three purposes: *training*, *validation*, and *testing*. The goal of the training is to find optimal weights (and other trainable variables) by minimising the cost function (Equation 3.3). Most machine learning models incorporate other tunable variables, *hyperparameters*. They are considered constants during the training process, but their values play a crucial role in the performance of the model. For example, the learning rate α in Equation (3.4) greatly affects on the success of the training process and thus model performance. The hyperparameters are optimised in the validation phase, usually relying on *grid-* or *random-search* [79] with a help of *k-fold cross-validation*. The general procedure of the cross-validation is as follows:

1. Shuffle the dataset to a random order.
2. Split the dataset into k groups.
3. For each group:
   (a) Hold out the group as the test data.
   (b) Fit a model with other groups and evaluate its performance.
   (c) Retain the evaluation score and discard the model.
4. Average evaluation scores to get overall performance for the whole dataset.

The grid-search involves defining a search space, a set of values for each hyperparameter to be evaluated, and applying the k-fold cross-validation for all possible combinations of the defined hyperparameter values. The random-search selects the performed trials randomly.

The purpose of the testing is to evaluate the model performance with unprecedented samples. A conventional procedure is to choose a fraction of samples to the test set randomly. Model performance is evaluated then with the test set, and the rest of the samples are used in the training phase. Sometimes, especially if the labeled data is an exiguous resource, the k-fold cross-validation is also used to obtain the final test results. That yields, however, circular reasoning as the same data is employed both in the validation and test phase.

A number of issues lurk in the randomised selection of the test set. A temporal, spatial, hierarchical, or phylogenetic structure may yield an *autocorrelation*, meaning that datapoints are depending on each other. In such conditions, the model may 'memorise' the linkage and provide overly optimistic results. Even more demanding, weather and climate data are also spatially autocorrelated, as the data have an internal geospatial structure. The temporal autocorrelation can be confronted by selecting the test set as a continuous-time series or picking a specific moment of the period, such as a day of the week. To minimise the effects of spatial autocorrelation, one needs to select samples from representing areas. Specific strategies to cope with the autocorrelation are analysed in [80].

However, these strategies do not come without repercussions. In impact forecasting, a sufficient amount of adverse weather is required to evaluate the results reliably. It is often hard to select a continuous time series, or long-enough moment of the period, having enough representative weather events without misspending the available training data. Thus, the resulting test data may not contain enough relevant samples for the reliable evaluation, or it may drain too many relevant samples from the training data resulting in improper training of the model. The issue is discussed more in Chapter 6.

Since extreme events are, by definition, rare, another challenge in impact forecasting is imbalanced data. Several different approaches exist to cope with the imbalance. One option is to give more value for the accuracy of minor classes, for example, by modifying the cost function [81], defined in Equation (3.3). The imbalance of the training set can also be removed by over-sampling or under-sampling [81]. The over-sampling creates new samples into the minority classes, while the under-sampling selects a subset of samples from the majority class. Naturally, one can combine these methods to create a hybrid approach. It is also notable that, for example, the ensemble methods, such as boosting and RF (presented in Sections 3.5.4 and 3.5.5), are not particularly fragile to the imbalance [82].

3.5 Summary of Relevant Machine Learning Methods

The following section provides a summary of popular machine learning methods applied in the tasks of power outage prediction in Chapter 4 and train delay predictions in Chapter 5. Since the list of available and experimented methods is enormous, the aim is not to be comprehensive but to provide a context for applied methods in this thesis. For a detailed description of the methods, the reader is asked to consult corresponding references.

3.5.1 Gaussian Naïve Bayes (GNB)

GNB [83] is a widely utilised method based on the Bayesian probability theory. The method assumes all samples to be i.i.d. This, as noted in Section 3.1, does not hold for the weather data but does not prevent the method from providing good results (e.g. [78, 24, 77]). GNB classifies the samples based on the rule:

$$\hat{y} = \arg\max_{y} P(y) \prod_{i=1}^{n} P(x_i \mid y),$$

where $P(y)$ is a frequency of class y and $P(x_i \mid y)$ is a likelihood of the ith feature, assumed to be Gaussian.

Since the samples are assumed to be i.i.d., each likelihood can be considered separately, which helps to cope with high-dimensional and small datasets. The likelihood estimations can be, nevertheless, conducted effectively and iteratively, which provides good scalability to large datasets. The main weakness of the basic form of GNB is a lack of expression power in more complex domains.

3.5.2 Generalised Linear Model (GLM)

GLMs generalise machine learning methods empowering the usage of labels that do not follow the normal distribution. Here, a particular extension of LR is considered. The GLM consists of three parts: 1) The labels are assumed to have a particular distribution with mean μ. 2) A LR, defined in Equation 3.1, is conducted. 3) The prediction is linked to the expected value $\mathbb{E}(Y|X)$ with a link function g so that $\mathbb{E}(Y|X) = \mu = g^{-1}(\mathbf{y})$ [84]. The method may be trained with the maximum likelihood estimation (MLE) or GD method. GLMs are applied to the weather data, for example, in the power outage predictions [85, 86].

3.5.3 Decision Trees (DT)

DTs are simple but yet powerful machine learning algorithms. An illustrative example of DT is shown in Figure 3.1. DTs are formed by *branches*

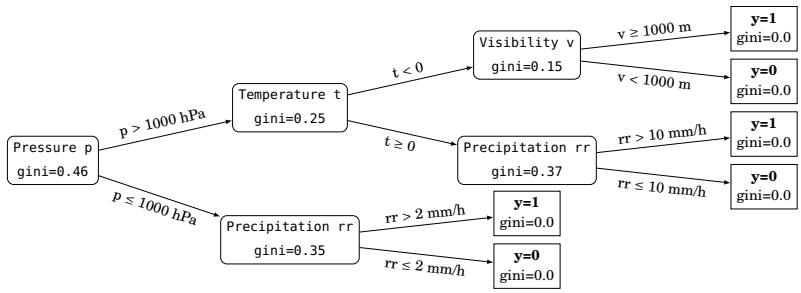

Figure 3.1. Example decision tree. Considered features and the corresponding thresholds in each node are deducted in the training process by minimising the cost function. This imaginary example use Gini-impurity.

with several *decision nodes* and one *leaf node*. Each decision node contains a logical 'if-then decision' regarding one predictive feature, and the leaf nodes are associated with labels. The decision criteria at each node, the best possible split, are determined during the model training by minimising a cost function, such as Gini-impurity [87]. The tree is then constructed by splitting the nodes recursively, starting from the root node and ending when a stopping criterion is satisfied or a maximum depth is reached. The training may be further improved, for example, by *pruning* the trees and arranging the nodes in the tree based on the feature importances [88]. DTs have been used, for example, to predict weather-related power outages [85, 86, 89].

3.5.4 Random Forest (RF)

RF builds on a random ensemble of DTs, discussed in Section 3.5.3. It aggregates the final estimate by averaging results of individual trees, constructed in four phases: (1) employ bootstrapping in order to generate a random sample of data, (2) randomly select a subset of features at each node in the tree, (3) determine the best split of features at the node based on selected loss function, and (4) grow the full tree [90]. RF is capable of coping with high-dimensional and imbalanced data [82]. The method is prone to overfit, which makes the hyperparameter tuning very important. Random forest classifiers (RFC) are widely applied in various weather applications [91]. They are also a popular choice to predict the weather-related power outages [92, 89, 93] and train delays [46].

3.5.5 Gradien Boosting Trees (GBT)

GBT resembles RFC, discussed in Section 3.5.4, in the sense that it exploits several randomly selected DT (see Section 3.5.3) to create the final prediction. Instead of averaging the results of individual trees, GBT sums the result of each tree multiplied by a learning rate α. [94] GBT has famous

variants: AdaBoost [95] is a GBT with exponential loss function and without cost-complexity pruning. XGBoost [96] use regression trees containing continuous scoring of each leaf in a tree and decide the final prediction based on the sum of the scores. The training of XGBoost considers also a cost complexity pruning regularisation, defined as $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda||w||^2$, where γ and α are hyperparameters of the model, T is a number of leaves in the tree, and w is leaf weights. The method has been used with the weather data, for example, in the context of outage prediction [89]. This thesis denotes gradient boosting tree classifiers as GBC and gradient boosting tree regression as GBR.

3.5.6 Multilayer Perceptron (MLP)

MLPs [97] represent artificial neural networks in the basic form. A simple feed-forward deep neural network contain multiple linear regression layers (defined in Equation 3.1) augmented with a bias term. The layers are connected with an activation functions $g(z) \rightarrow s$. Each node i at the layer l gets output as:

$$s_{l,i} = g\big(\sum_{i=0}^{N} w_{l,i} s_{l-1} + b_{l,i}\big),$$

where $w_{l,i}$ is a weight variable, $b_{l,i}$ is a bias variable and N is a number of nodes at the layer l.

The training of MLPs is conducted using GD in the process called back-propagation. Overfitting is typically controlled by dropping connections randomly during the training process and stopping the training process before the model gets overfitted. Figure 3.2 illustrates the MLP. Deep neural networks are able to learn a representation of the input at their hidden layers, which makes them very adaptive and powerful methods. The non-linear activation functions and possibly several hidden layers enable MLPs to learn nonlinear learning problems. As a downside, MLPs have a large number of hyperparameters, and the correct network architecture requires careful design from the developer. They also require a large dataset to be appropriately trained.

3.5.7 Support Vector Machines (SVM)

SVMs [98] employ typically nonlinear kernels to form high-dimensional feature space and maximise the distance between training samples and the hyper-plane. The kernels are used to reform a nonlinear classification problem to a linear one as illustrated in Figure 3.3. Popular kernels are for example Gaussian radial basis function (RBF) [98] and a dot product kernel [99]. Being able to operate in the high-dimensional feature space without additional computational complexity is one of the main advantages of SVM, especially for high-dimensional datasets. Moreover, the kernels

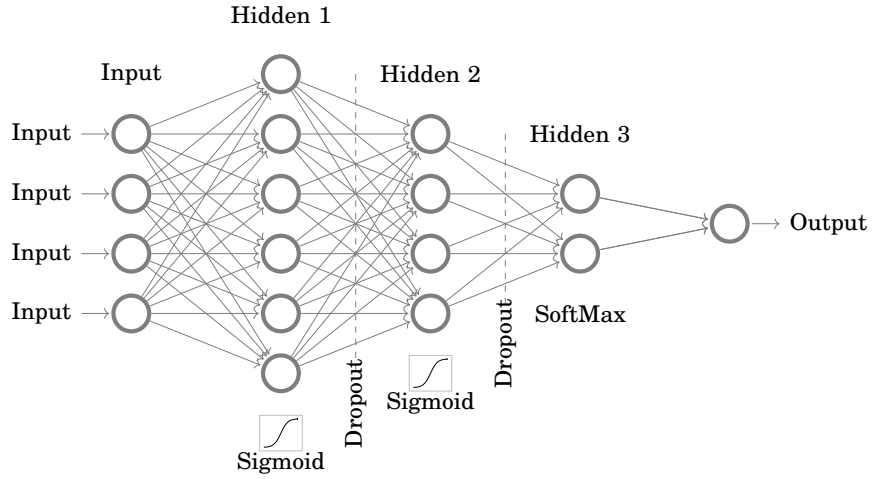

Figure 3.2. Example multilayer perceptron. The nodes are shown as circles and weights as lines between the nodes. This example has Sigmoid activation function ($\frac{1}{1+e^{-z}}$) at the first two hidden layers and SoftMax layer, producing categorial output at the last hidden layer.

provide a natural way to capitalise on domain-specific expert knowledge. SVM training is a convex optimisation problem without a local minimum. The computational complexity of the training process is, however, between $\mathcal{O}(n^2)$ and $\mathcal{O}(n^3)$ [100] and some kernels may turn that into a very memory-intensive process as well.

3.5.8 Gaussian Processes (GP)

GP interprets the observed datapoints as a realisations of a Gaussian random process. GP can be seen as a stochastic process defining a Gaussian distribution over functions:

$$f(\mathbf{x}) \sim \mathrm{GP}(\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}\mathbf{x}'}),$$

where

$$\mu_{\mathbf{x}} = \mathbb{E}\big(f(\mathbf{x})\big),$$
$$\Sigma_{\mathbf{x}\mathbf{x}'} = \mathbb{E}\big(\big(f(\mathbf{x}) - \mu_{\mathbf{x}}\big)\big(f(\mathbf{x}') - \mu_{\mathbf{x}'}\big)\big)$$

In other words, GP model the underlying distribution of dataset $\mathbf{X}$ together with labels $\mathbf{y}$ as a multivariate normal distribution. The joint probability distribution $p(\mathbf{X},\mathbf{y})$, assumed to follow Gaussian distribution, thus covers the space of possible values for the predicted function. The distribution $p(\mathbf{X},\mathbf{y})$ is updated using Bayesian inference, and the posterior distribution is fitted by maximising a log-marginal-likelihood. A prior for mean $\mu_{\mathbf{x}}$ is typically assumed as either zero or mean value of the data. The covariance matrix $\Sigma_{\mathbf{x}\mathbf{x}'}$ is generated with *a kernel*, a pairwise covariance

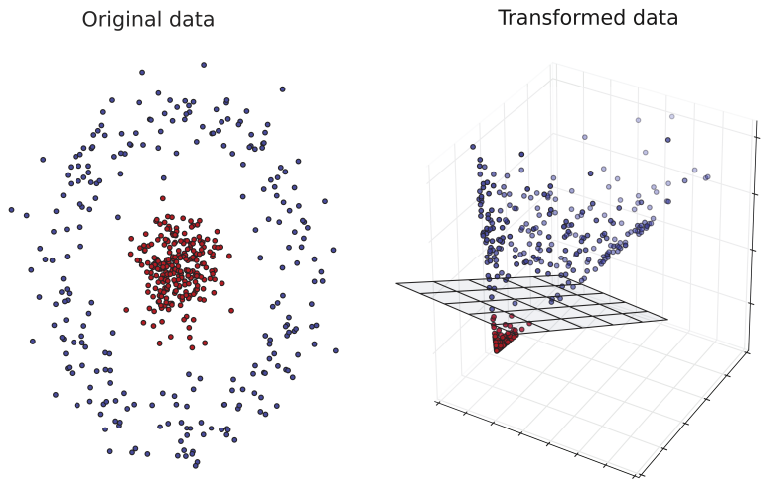

Figure 3.3. Kernel trick illustrated. (a) visualises a linearly non-separable classification problem. In (b) the same data is transformed with kernel $\Phi(\mathbf{x}) = (\sqrt{2}x_1x_2, x_1^2, x_2^2)$. The decision boundary of SVC classifier is presented with gray surface.

function on all points, representing a pairwise interdependence of the datapoints.

GP is an extremely flexible and powerful but computationally complex method. Similarly to SVM in Section 3.5.7, GP exploits the kernel function, which opens the developer an avenue to effectuate the domain-specific knowledge. Due to the Bayesian inference, the model provides probabilistic predictions. The complexity casts, however, challenges to utilise large datasets and high-dimensional data. A detailed view on the GP is given, for example, in [99].

4. Predicting Power Outages

As discussed in Chapter 1, the power outages caused by extreme weather events are the key problem for power grid operators. With an accurate outage prediction, the power grid operators may prepare to repair the damages with the right workforce level. The predictions may also help the operators analyse what has happened and target the needed resources to the damaged regions. Finally, they can communicate anticipated disturbances to the clients beforehand. While some applications for preparedness already exist (e.g., [101]), they are relatively uncustomary. For example, in Finland, the power grid operators use weather information, warnings, and advisories made by duty forecasters in anticipation.

The chapter starts with a review of the already conducted research. The proposed methods for convective and extratropical storms are presented in Sections 4.2.1 and 4.2.2 correspondingly. Both sections describe the method and present the key results. Finally, the chapter ends with a section 4.3, discussing delivering the information to the end-users.

4.1 Existing Work

The power grid network started to evolve in the US in the late 1880s [102] and already in 1939, a relationship of power outages and extreme weather was studied [103]. Early methods simulate the power grid system behaviour under extreme weather, typically with the Markov Processes [104, 105, 106, 107, 108, 109]. Until the 1990s, the weather was usually classified in two or more categories based on its damage potential. However, as the accuracy of the weather forecasts improved and the power grids evolved, the power outage models required more granularity. Therefore, several applications started to employ individual weather parameters, such as wind and temperature, in the power outage predictions using Monte Carlo simulation [110, 111, 112, 113, 114] and heuristic model [115]. Icing was studied using numerical simulation [116] and statistical methods [117, 118]. A yearly weather-related failure analysis using Poisson regression

and a Bayesian network model was implemented in [119].

After the 2010s, the research agenda has turned more into machine learning algorithms. The studies can be roughly categorised into three classes based on the weather phenomenon size and intensity: small-scale convective storms, extratropical storms, and tropical cyclones. The majority of the studies address the hurricane-induced power blackouts in northern America (i.e.[120, 121, 122, 123, 124]). Typical methods use DT (see Section 3.5.3), RF (Section 3.5.4), GLM (Section 3.5.2), Poisson regression, and fuzzy logic to estimate blackouts. This thesis focuses on smaller-scale events and thus omits a more detailed analysis of the hurricane-caused blackouts.

4.1.1 Convective Storms

In the early 2000s, only a few studies addressed local and ephemeral weather phenomenons, such as thunderstorms [125, 110, 126, 127]. Improved data processing capabilities, better weather forecasting, and advances in machine learning techniques, discussed in Chapters 2 and 3 respectively, have opened an avenue for several new power outage methods. Many of the new methods are focusing on convective storms. Authors of [128] classified the weather by temperature, wind speed, and amount of lightning to forecast cumulative power outages applying an empirical exponential function over time. Kankanala et al. applied LR [129] (see section 3.2), MLP [130] (Section 3.5.6), and AdaBoost [131] (Section 3.5.5) with ground observations to predict weather-caused outages. Li et al. introduced an outage prediction method [85] which was later modified to employ power grid topology as well [86]. The methods in [85] and [86] applied DT (see Section 3.5.3) and GLM (Section 3.5.2) to the predefined regions. Yue et al. introduce a Bayesian outage probability prediction that utilises weather radar data from several sources combined to the geographically unified grid [132].

Wanik et al. evaluated DT, RF, GBT (see Section 3.5.5), and an ensemble decision tree regression (ENS) with NWP data (discussed more in Section 2.2) to predict several types of storms, including both thunderstorms and extratropical storms [89]. Similarly, quantile regression forests (QRF) and Bayesian additive regression trees (BART) were used with NWP data in [133]. Shield et al. studied an outage prediction by employing RFC to NWP data [93].

4.1.2 Extratropical Storms

Relatively few former works focus on outages caused by extratropical storms. Some methods employing QRF, BART, and RFC, already discussed in the context of convective storms (Section 4.1.1), employ NWP data and

have therefore a utility in predicting the outages caused by extratropical storms [89, 133, 93]. These studies have been extended in [92], where the authors build on [89] and [133] by classifying the storms by their types (thunderstorms, extratropical storms, hurricanes) and using different calibrations for meteorologically distinct storms. The authors of [134] continued this work by adding uncertainty information to the predictions.

4.2 Proposed Method to Predict Power Outages

This section proposes a novel method to predict power outages caused by convective storms and extratropical storms.

4.2.1 Convective Storms

The proposed method to predict the power outages caused by convective storms is first presented in Publication I and then extended with more detailed analysis and evaluation in Publication II. The method is based on the object-oriented approach and utilises the storm cell identification and tracking developed in [25]. The storm objects are first identified from the radar reflectivity CAPPI images with a solid 35 dBZ contour threshold. The particular threshold is selected to detect complete storm systems like multicellular storms [135, 55]. Section 2.1 describes the data in more detail.

After the identification, GDBSCAN clustering [136] is applied to the storm objects. The particular clustering algorithm is chosen as a similar approach has provided prominent results in the previous applications [137, 138]. As an advantage, it can discover an unspecified number of clusters of any shape. In the clustering, a storm object is considered as a *core object* C if the cumulative area of its nearby storm objects inside *a radius* exceeds *an area threshold*. The objects not exceeding the area threshold are considered as *outliers* O. Objects are said to be *directly reachable* if they are inside the radius from a core object. Objects can be directly reachable only from core objects. Objects are considered to be *connected* if a path $p_1, ..., p_n$ exists from C to O so that p_{i+1} is directly reachable from p_i, $p_1 = C$, and $p_n = O$. The definition of connection implies that all objects along the path except p_n need to be core objects. All connected objects form a *cluster* and the objects outside the radius and not filling the area threshold to be core objects are considered noise. This method used 20 km^2 as the area limit, and 2 km as the radius based on the evaluation in [137].

The formed clusters are then tracked, and their movement is predicted based on the idea introduced in [139]. The previous time step clusters are interpolated to the current time step employing optical flow [64]. If the

interpolated cluster overlaps with the current time step cluster, they are considered to be connected. The nowcast is conducted 2 hours ahead with a 5 minute time interval using Kalman filtering as in [137].

The clustered storm cells are then classified into 4 categories based on their potential to induce outages. The outage data is obtained from two Finnish power grid companies, shown in Figure 4.2 a, and contains in total 33 858 outages. Each outage is reported as a malfunctioning transformer (node in the power grid) near the actual damage location. The storms are cast into the classes based on the share of the malfunctioning transformers of all transformers covered by the storm object. The classes are introduced in Table 4.1. They are designed to provide a simple and quickly interpretable view to power grid operators. Detailed description and analysis of the outage data are presented in Publication III.

Table 4.1. Storm cell class definitions

Class	Share of transformers
0	no damage
1	0 - 10 %
2	10 - 50 %
3	50 - 100 %

The classification is conducted based on the storm features derived from the clusters themselves (such as area, age, and location), the weather radar measurements (such as reflectivity dBZ described in Section 2.1), the supporting features from ground observation stations, and thunder strikes detected by the lightning detection network. That is to say, the training data consists of $\mathbb{X} = \left(\mathbf{x}^{(i)}, y^{(i)}\right) \in \mathbb{R}^{N \times D}$, where $\mathbf{x}^{(i)}$ represents characterising features of the cluster at one time step, $y^{(i)}$ represents a the storm class at the corresponding time step, N is a dataset size and D is a number of features in $\mathbf{x}$. Publication I employed in total 559 071 samples and Publication II employed in total 885 976 samples. Both papers employed 16 features.

RFC, discussed in Section 3.5.4, and MLP, discussed in Section 3.5.6, were experimented in Publication I and Publication II. Gini-impurity [87] and cross-entropy (Equation 3.2), were used to train the RFC and MLP respectively. Most storms are not causing damage to the power grid, and therefore the imbalance of the dataset is a pivotal issue, as noted also in [140]. As already discussed in Section 3.4, the imbalance can be coped with several ways. For example, [141] and [93] propose a two-stage prediction, with the first stage predicting whether any damage will occur and the second stage predicting the amount of damage. This work equalised class sizes by over-sampling using the synthetic minority over-sampling technique (SMOTE) [142].

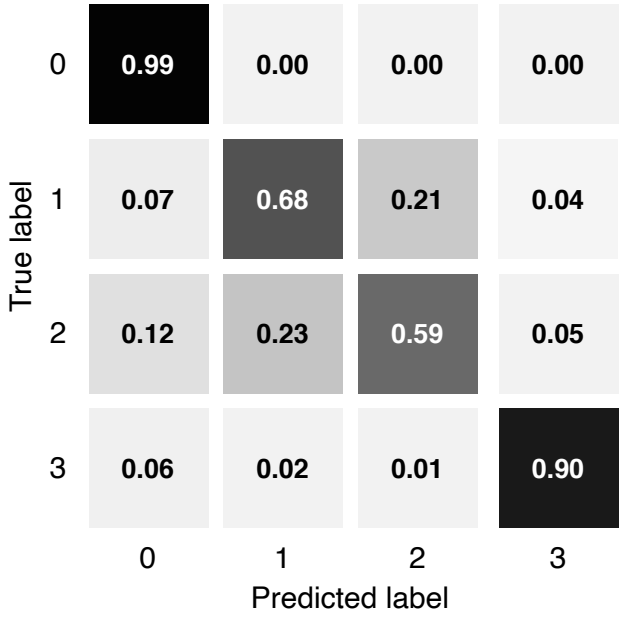

Figure 4.1. Normalised confusion matrix of RFC. Each cell represents the probability of predicted and true label combination. For example the upper left cell tells a probability that the model predicts class 0 when true class is 0.

Results

RFC provided better results than MLP. A macro average of F1-score, defined in Equation 4.1, was selected as the key metrics. The particular metric gets values from 0 (worst) to 1 (best) and was chosen to represent a harmonic mean of recall and precision with one number. The test set included 25 percent of the total samples selected randomly.

F1-score:

$$F1_{macro} = \frac{1}{N}\sum_{\lambda=1}^{N}\left(\frac{precision_\lambda \times recall_\lambda}{precision_\lambda + recall_\lambda}\right), \tag{4.1}$$

where

$$precision_\lambda = \frac{t_{p\lambda}}{t_{p\lambda} + f_{p\lambda}},$$

and

$$recall_\lambda = \frac{t_{p\lambda}}{t_{p\lambda} + f_{n\lambda}},$$

and $t_{p\lambda}$ is amount of true positive samples, $f_{p\lambda}$ false positives and $f_{n\lambda}$ false negatives in class λ.

RFC gained 0.70 as a macro average which can be considered as an excellent result for the task. The results can also be illustrated with a confusion matrix depicted in Figure 4.1. Each cell in the matrix shows the probability of predicted and true class combination. The accuracy is

excellent in the class 0 (no damage) and the class 3 (high damage) but lower in the middle classes.

While the comprehensive comparison between the results of previous studies and this thesis is an extremely hard or impossible task due to the differences in data, environment, and exact task definition, the attained results are auspicious. The object-oriented approach equips the model to focus on the relevant data and provide important information about the storm movement and life cycle that would be complicated to obtain else. The advantages and disadvantages of the object-oriented approach are discussed more in Chapter 6.

The proposed solution has some significant disadvantages as well. It does not exploit 3-dimensional information. The 3-dimensional data could be used with more advanced object identification and tracking algorithm, discussed in Section 2.1. Some significant features are also missing. Forest information, used in Publication III, and information about ground frost and tree leaves would most probably improve the results.

4.2.2 Extratropical Storms

Extratropical storms, distinguishable from convective storms and tornadoes in time-span, geographical scale, characterising meteorological parameters, and applicable methods for monitoring and tracking them, as described in Section 2.2, are responsible for over half of all losses related to the natural hazards in Europe [143]. The method presented in section 4.2.1 provides a crucial tool in mitigating the effects of convective storms but has a limited capacity with extratropical storms. Hence, a specific tool for predicting the outages caused by these storms is required. In addition to the more accurate prediction, the dedicated tool provides power grid operators a pivotal advantage to extend the prediction lead time to days, as the forecast is based on the NWP data ranging days ahead.

All existing machine learning-based power outage prediction methods, discussed in Section 4.1, have been built on regular grid or predefined areas. Publication III introduces a novel approach, where the object-oriented power outage prediction, presented in Section 4.2.1, is modified to operate with an analysis and NWP data, discussed in Section 2.2. The overall process is the following: (1) identify potentially devastating storm objects, (2) track the identified objects, (3) extract characterising features of the objects, and (4) classify the objects based on their damage potential.

The process starts with an identification of wind and pressure objects. The wind objects are identified from the grid by finding contour lines of 10 metre maximum wind gust fields and the pressure objects from the surface pressure fields using single thresholds. Thus, the identification process yields solid polygons with wind gust above the threshold and pressure polygons below the threshold. The wind objects are considered

and called from here on as the storm objects. The pressure objects connect the potentially distant storm objects around a low-pressure center to a single storm event and thus enhance the tracking process. A threshold of 15 ms^{-1} is selected for the wind gusts and 1000 hPa for the pressure objects since several studies indicate that in Finland the forest damages start to occur above 15 ms^{-1} winds [144, 145].

The second phase is to track the objects so that characterising features – such as movement, size, and age – from the whole life cycle of the storm can be extracted. Each object is first connected to the nearest pressure objects within *a distance threshold* from the current and preceding time steps. If the pressure objects are not found, the storm object is connected first to the nearest storm object within the distance threshold and second to the nearest storm object on the preceding time step. The preceding time step objects are required to be inside *pressure and storm object movement thresholds*. The detailed algorithm is presented in Publication III.

Two main reasons necessitate a different tracking algorithm than with the convective storm cells presented in Section 4.2.1. First, DBSCAN does not find proper clusters for the extratropical storm objects with large solid polygons, possibly far away from each other. Second, the optical flow algorithm cannot consider the existence of the pressure objects; the advantage of the bespoke tracking algorithm is the ability to consider all storm objects around one low-pressure object as a single storm event without merging them to a single polygon.

The distance threshold between the storm and the pressure objects was set to 500 kilometres. Since extratropical storms are typically approximately 1000 kilometres wide [72], the devastating storm objects are expected to locate no further than 500 kilometres from the pressure objects. The movement threshold is 200 kmh^{-1} for the storm objects and 45 kmh^{-1} for the pressure objects. That is to say, the storm objects are assumed to move at a maximum of 200 kilometres and the pressure objects at a maximum of 45 kilometres in an hour [72]. Convective storms sometimes move faster but are outside the focus of this method.

The third phase of the process is to extract characterising features of the storm objects. Three groups of features are used. The first group consists of features describing the objects themselves. The object size is the most important feature, while the object movement direction and speed also contribute to the prediction. The second group contains weather conditions such as wind speed, temperature, and other variables. The values are aggregated as a minimum, maximum, average, and standard deviation calculated over the object coverage. The third group carries forest information such as tree height, age, et cetera. Similarly to the weather parameters, the values are aggregated over the object coverage. Finally, including the number of outages, is used in the training phase as a label. In total, 35 parameters, listed in Publication III, are used in the

classification.

The fourth phase is to classify the objects based on their damage potential. Table 4.2 lists the classes. As in Section 4.2.1, the classes were designed in collaboration with power grid companies and aimed to provide a quick view of the situation. Class 1 is chosen to represent 80 percent of all devastating storm objects. As in Section 4.2.1, most storms are not causing any damage and the classes are highly imbalanced.

Since extratropical storms are larger and longer-lasting events than convective storms, lower geospatial and temporal accuracy are required in the classification. Lower requirements opened the author a possibility to use a larger, national outage dataset obtained from most Finnish power grid companies and containing the whole of Finland. Instead of accurate location, the outages are only reported to be in one of the five different regions, shown in Figure 4.2 b. Publication III compares the results between classifiers trained with the local dataset, used in predicting convective storms in Section 4.2.1, and the national dataset. The national dataset provided better results.

Table 4.2. Class definitions in the task of predicting extratropical storms-related power outages

Class	Outage limit	Dataset size	Class description
0	0	76 215	no damage
1	1- 140	14 417	low damage
2	≥ 141	3 085	high damage

Publication III compares the performance of RFC (Section 3.5.4), SVC (Section 3.5.7), GNB (Section 3.5.1), GP (3.5.8), and MLP Section (3.5.6) for the task. As discussed below, SVC achieved the best F1-score and most harmonic classification results in the evaluation. Similar to [141] and [93], the classification is conducted in the phases: First, class 0 (no outages) and other objects are distinguished using SVC with radial basis function (RBF), defined in Equation 4.2. Second, the objects with predicted outages are divided into classes 1 and 2 employing SVC with a dot-product kernel, defined in Equation 4.3 [99]. The approach resembles the widely-used one-vs-one classification scheme, where a binary classifier is applied for each pair of classes. Contradicting to the one-vs-one scheme, the model presented here employs separate kernels for separate pairs. The class imbalance, discussed more in Section 4.2.1, was coped with SMOTE.

RBF kernel:

$$k_{RBF}(\mathbf{x},\mathbf{x}') = \exp\left(-\gamma||\mathbf{x}-\mathbf{x}'||^2\right) \tag{4.2}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are two samples in the input space and γ is a kernel coefficient parameter.

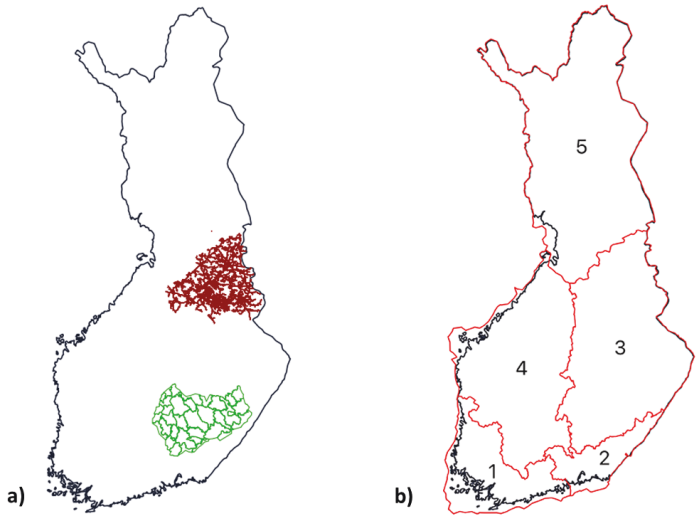

Figure 4.2. a) Geographical coverage of the local outage data (used in Section 4.2.1). The power grid of one company is illustrated with red lines, and the operative areas of another company with green lines. b) National outage dataset regions. Outages are obtained from most power grid companies in Finland and aggregated to the regions shown in the figure. The figure is originally published in Publication III (license CC4BY).

Dot-product kernel:

$$k.(\mathbf{x},\mathbf{x}') = \sigma_0 + \mathbf{x}\cdot\mathbf{x}' \tag{4.3}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are two samples in the input space and σ is a kernel inhomogenity parameter.

Results

As with convective storms in Section 4.2.1, the F1 macro average (Equation 4.1) was used as the primary metric to evaluate the performance. Because of the autocorrelation issue, presented in Section 3.4, the evaluation was done with both a randomly selected test set and a continuous period test set. SVC gained 0.60 as a macro average of F1 for both randomly selected and continuous test sets. The confusion matrix produced with the randomly selected test set is shown in Figure 4.3.

The gained performance is significantly lower than with convective storms, discussed in Section 4.2.1. The possible reasons for the lower performance are the coarser spatial and temporal resolution of the data. While destructive parts of the convective storms can be relatively reliably detected with 250 metres spatial and 5 minutes temporal resolution, extratropical storms are monitored with several kilometre accuracy. The precise devastating locations of the wind gusts are not obviously detectable from the data. The convective storms are also always occurring in summer when trees have leaves and the ground is unfrozen. The extratropical storms may occur any time of the year, but the presented method is missing information about the leaves and the ground frost, significantly affecting

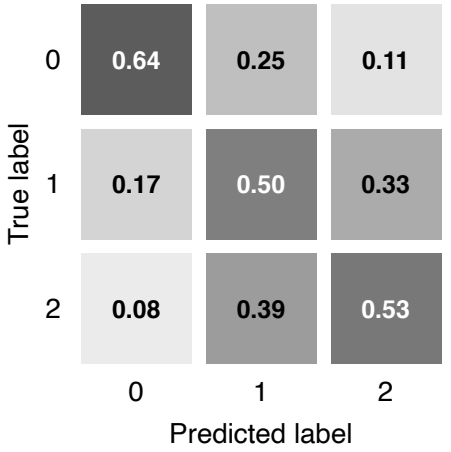

Figure 4.3. Normalised confusion matrix of SVC. Each cell represents the probability of predicted and true label combination. For example the upper left cell tells that the model predicts class 0 with 64 percent probability when true class is 0.

the impact of the storm. Unfortunately, such parameters were not easily available at the time of conducting the research.

The methods have been developed in collaboration with three Finnish power distribution companies[1] and are used by the grid operators. At the time of writing, the convective storm application provides evidence of being useful in the operations, and the experiments of the extratropical storm application have been started. Chapter 6 debates more about the results, their limitations, and the future directions.

4.3 User Interface for the Power Outage Predictions

Processing objects instead of grid or individual points opens an avenue for rich applications since the objects may be enriched with additional information and interactions. A bespoke web-based user interface (UI) was developed to provide the forecast to the power grid operators. Figure 4.4, shown already in Chapter 1 as an example of impact forecast, depicts the UI for the convective storm-related power outage prediction in Section 4.2.1. Figure 4.5 extends the UI to work with extratropical storms presented in Section 4.2.2. If convective storms are selected, the operator sees all identified and forecasted storm objects from the preceding and the following 30 minutes. The identified storm objects are marked with a solid colour based on the predicted object class. The forecasted objects are marked with a texture with the corresponding colour. If extratropical storms are selected, the operator sees only one time step at a time.

[1]Loiste sähkönsiirto ltd, Järvi-Suomen Energia ltd, and Imatra Seudun Sähkönsiirto ltd

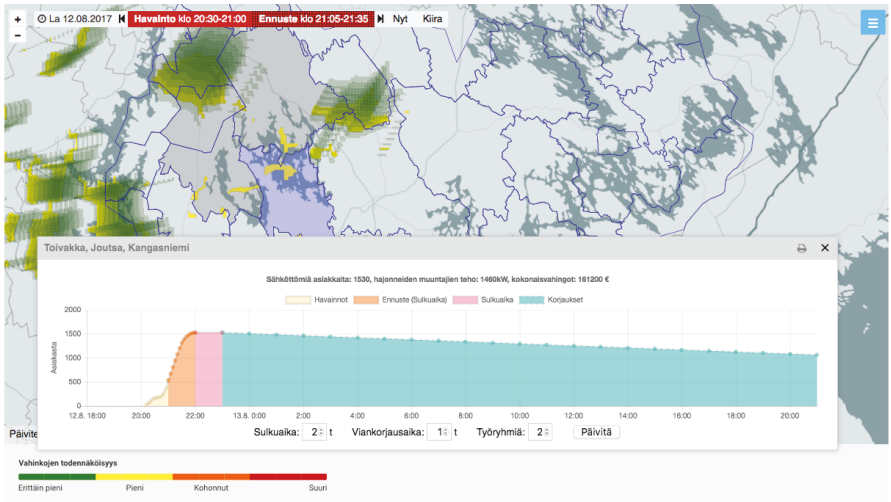

Figure 4.4. A user interface is providing the forecast of convective storms to power grid operators. The map view presents the categorised storm objects, and the graph estimates the number of households being disconnected from the power grid at the selected area as a function of time. Blue lines on the map represent operating areas, following roughly Finnish municipalities.

Finnish power grid companies are obligated to compensate customers for the interruptions in electricity distribution. Urban-area customers are eligible for the compensation after 6 hours, and rural area customers after 36 hours. The operator can use the UI to select municipalities of interest and get an estimated number of households without electricity along with derived liabilities for the power grid company. The estimation is shown in the graph over the map in Figure 4.4. The first part of the graph, shown with yellow, represents the occurred outages, while the orange part shows predicted outages. The area with pink colour stands for (constant) waiting time before restoration can begin. After the waiting time, the damages are assumed to be repaired at a constant speed depending on the number of teams. The operator can adjust the number of teams and estimated repairing time to modify the estimated liabilities.

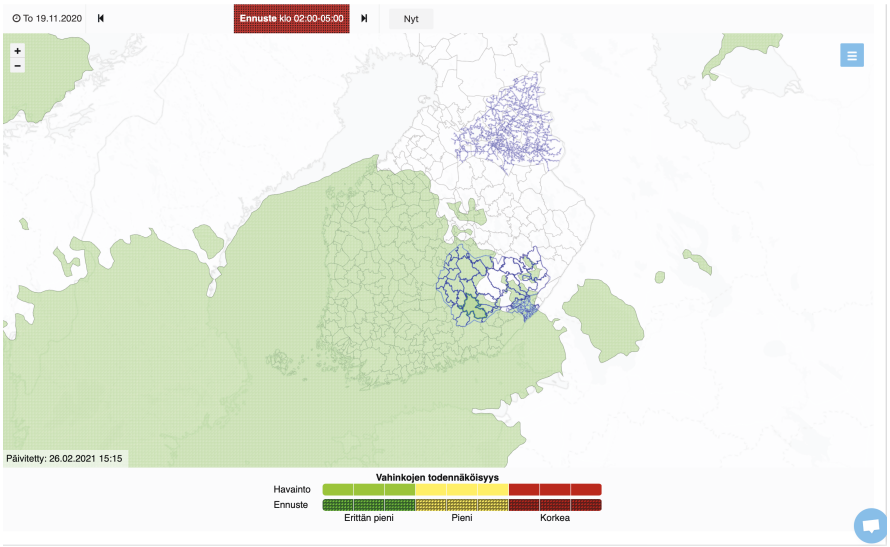

Figure 4.5. A user interface providing the forecast of extratropical storms to power grid operators. The map view presents the categorised storm objects. The user may select the area of interest and get similar impact estimation as shown in Figure 4.4.

5. Predicting Weather-Inflicted Train Delays

As discussed in Chapter 1, train delays cause considerable economic losses. The punctuality also affects crucially on the attractiveness of rail traffic [146, 147]. Predicting the train delays as early as possible is a key to reduce them. If anticipated well in advance, rail traffic operators may take several actions to mitigate the effects of the train delays: (1) They can reschedule cargo trains and cancel part of the trains to keep the rest better in time. (2) Especially with long-term predictions, rail traffic support functions such as resource planning may be better prepared. (3) Anticipated delays may be communicated to passengers. (4) In some countries, the rail traffic operators may also reroute the trains [148, 149]. Such predictions are also essential information for weather duty forecasters in issuing impact-based weather warnings.

While diverse reasons induce the delays, most severe rail traffic disruptions are inflicted by adverse weather. Especially snowfall and heavy wind, affect on journey times in northern Europe [150, 44, 41, 43], China [45, 151, 152], and Toronto [153]. During 2008 – 2010, 60 percent of the freight railway delays in Finland were caused by the winter weather [40]. Punctuality have been noticed to decrease exponentially when the temperature drops below 0 °C [154]. Brazil et al. found that the time of the year and prevailing weather are better predictors for the delays than the published schedules for the trains [155].

Despite the evident connection between the adverse weather and the delays, current methods employed for preparedness are typically rather manual and are based on the rail traffic interpretation of current weather forecasts and warnings. Moreover, although preparing for the mundane challenging conditions would also help to prepare ultimate extreme events, many operators are focusing on the natural hazards only [156]. More systematic preparedness may be categorised into four types: topological, simulation, optimisation, and data-driven [157].

This chapter describes a novel data-driven method to predict the weather-inflicted train delays days ahead. Instead of providing solely underlying weather information, the method shows an estimation of anticipated train

delays. Next, existing work is discussed in Section 5.1. After the literature review, the proposed method is introduced with key results in Section 5.2. Finally, delivering the results to the rail traffic operators is discussed in Section 5.3.

5.1 Existing Work

Various works exist to predict a short-term delay propagation in the rail network hours ahead (i.e., [158, 159, 160, 161, 162, 163, 164, 165]). Some works exploit also weather parameters, such as temperature, wind, precipitation, visibility and snow depth, in operations [166], railway upkeep [167, 168]. Serdar et al. utilised extreme wind, snowfall, fog, rainfall, flood, and temperature to create a Bayesian-network-based model of derailments in turnouts caused by extreme weather [169]. Molarius et al. created a macro-level weather risk indicator for all transport modes in [42], considering precipitation, blizzards, wind, and temperature in their study. Oneto et al. compared RFR, SVM, and MLP for the short-term prediction of the delays. The primary features of the prediction were train punctuality and schedule information of approximately 1 000 trains from 6 months, but the models were augmented with six weather parameters [46].

5.2 Proposed Method

While the weather parameters have been considered in the short-term delay propagation predictions ranging hours ahead, the author is not aware of any long-term delay prediction reaching days ahead. Publication IV introduces a novel delay prediction method based on the weather information. The method predicts the anticipated average delay experienced at a train station.

The delay prediction is formalised as a supervised learning problem with weather parameters as primary features and an average delay of passenger trains as labels (features and labels are discussed in Chapter 3). The average delay is defined as a mean of the delays of all trains arrived at a specific station during an hour. The delay is defined as the difference between nominal travelling time and the actual travelling time from the previous station.

The method contains two complementary phases: The first phase is to conduct a binary classification to predict whether the delays exceed a 10-minute threshold at a train station. The 10 minutes threshold is chosen based on personal communication with the Finnish rail traffic operators. It is set to distinguish broad disruptions from 'individual delays.' The threshold is not, however, a specific limit for the rail traffic operators

to take action. The second phase is a regression problem predicting the amount of delay as a continuous variable.

The delay prediction is based on 13 weather parameters, train station latitude and longitude, and a month obtained from the train arrival time. The features contain typical weather parameters, like temperature and wind, and also parameters such as cloud base, pressure, and visibility which are expected to indicate challenging weather parameters indirectly. Since the precipitation is an influential element of the challenging weather conditions [155], the accumulated precipitation for three and six hours was calculated. The month is used to distinguish different conditions during winters and summers [155] and the train station location to differentiate circumstances at northern and southern stations. Publication IV lists the employed features and discusses them in more detail. The prediction is based primarily on the weather parameters and excludes the effects of the train network topology and operations since the propagation of ongoing delays has only minimal or no effect in the delays of the following days [170]. Challenging weather also affects broad areas [72] and many adjacent train stations concurrently in a geographically smooth country as Finland, so the rail network topology is assumed to have only a limited effect on the prediction performance. The model may, nevertheless, recognise challenging locations based on the train station location.

The weather observations were obtained from the corresponding one-hour time slot with the average delay within a 100-kilometres radius from each train station. They are aggregated as the minimum, maximum, and mean for each parameter. The 100-kilometre range covers whole railway sections between adjacent stops in virtually all records in our data. The observations were gathered around the train stations instead of railway sections since the delays are considered as the average delay of trains arriving at a station from all directions.

Publication IV compares GNB, RFC, GBC, LR, RFR, and GBR, described in Section 3.5, in the task. As already mentioned, the delays may naturally occur because of many reasons. As shown in Publication IV, a great majority of the recorded delays are without any reason code. The models were trained with both weather-related delays and all delays. The weather-related dataset contains only records marked with a ‘weather’ reason code. The whole dataset contains all records except the ones marked with a reason code not relating to the weather. Table 5.1 lists the characteristics of both datasets.

The data is naturally imbalanced as most of the trains are running in time. Section 3.4 presents different strategies to cope with the imbalance. Publication IV analyses different over- and under-sampling ratios in the task of train delay prediction. The best results with GNB were obtained by the under-sampling, and with RFC and GBC by over-sampling the minority classes using SMOTE [142], as in power outage prediction in Chapter 4.

Table 5.1. Characteristics of train delay data divided to two different dataset: the weather-related dataset and the whole dataset.

	Weather-related	Whole dataset
Row count	7 350	10 679 977
Number of train stations	56	486
Share of class 0 rows	67 %	92 %
Share of class 1 rows	33 %	8 %

5.2.1 Results

As with the power outage prediction, discussed in Chapter 4, the primary metrics in the classification evaluation was the macro-average of F1-score, defined in Equation (4.1), and the confusion matrices. The methods were evaluated both with a randomly chosen dataset and a continuous time series to cope with the autocorrelation issue, discussed in Chapter 3. The most challenging month from the whole data, February 2011, containing long-lasting severe delay periods, was selected as the continuous test set. The randomly chosen test set may be seen as a 'normal situation' and the continuous test set as an 'exceptional situation.'

Publication IV reveals the detailed metrics. GBC performed best in the task, with the F1 macro average ranging from 0.53 to 0.72 depending on the dataset and the test set selection method. Figure 5.1 shows the confusion matrices. The model gains adequate performance for the randomly selected test sets, representing 'normal situation' but tends to be biased for the continuous test sets, representing 'exceptional situation.'

The regression was evaluated using mean absolute error (MAE), root mean square error (RMSE) and Brier Skill Score (BSS) defined as:

$$BSS = 1 - \frac{RMSE}{RMSE_{ref}},$$

where $RMSE$ denotes root mean square error of the results and $RMSE_{ref}$ denotes a root mean square error calculated with a mean value over the dataset.

RFR provided the best results in terms of the metrics achieving 2.82/4.63 as MAE, 6.64/7.06 as RMSE, and 0.06/0.26 as BSS for the randomly chosen and the continuous test set of the whole dataset correspondingly. Errors were larger for the weather-related test set: RFR achieved 8.40/9.59 as MAE, 14.86/12.71 as RMSE, and 0.15/0.06 as BSS for the randomly chosen and the continuous test set correspondingly. Figure 5.2 portrays the occurred and predicted average delays over train stations in February 2011. The figure reveals that even though GBR did not provide as good metrics as RFR, it provides very similar predictions. The larger errors derive from GPR's tendency to predict longer delays, increasing the errors

Figure 5.1. Normalised confusion matrices of GBC for both the whole and weather-related dataset and the continuous and randomly selected test set. Each cell represents the probability of predicted and true label combination.

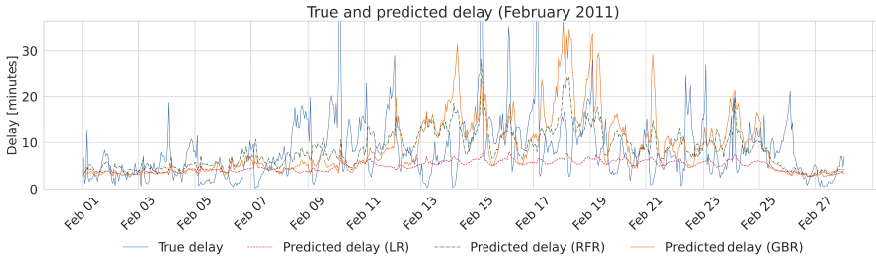

Figure 5.2. Actual and predicted average delay over train stations in February 2011 (test set of the whole dataset). Figure from Publication IV.

if the prediction is wrong or wrongly timed.

Surprisingly, the models trained with the whole data, including also delays not related to the weather conditions, provided similar or even better results than those trained with only the weather-related delays. Two reasons are assumed to explain the effect: The weather-related dataset is relatively small for the sufficient training of the models. On the other hand, as discussed in Publication IV, a remarkable share of delays in the whole dataset is evidently related to the weather even though stored without any reason codes.

The results provide evidence that the proposed method can predict train delays with high accuracy if trained with sufficient history of good quality data. The results are in line with the previous studies [155, 46]. The consequences and the limitations of the results are discussed more in Chapter 6.

5.3 User Interface for the Train Delay Predictions

The train delay predictions were delivered to the rail traffic operators via a web-based UI shown in Figure 5.3. The train stations are shown on the map coloured based on the predicted class. The user may select the interest of time with a time slider and select a train station to get the predicted amount of delay, shown as a graph under the map. Associated weather parameters, such as precipitation, may also be visualised as a map overlay.

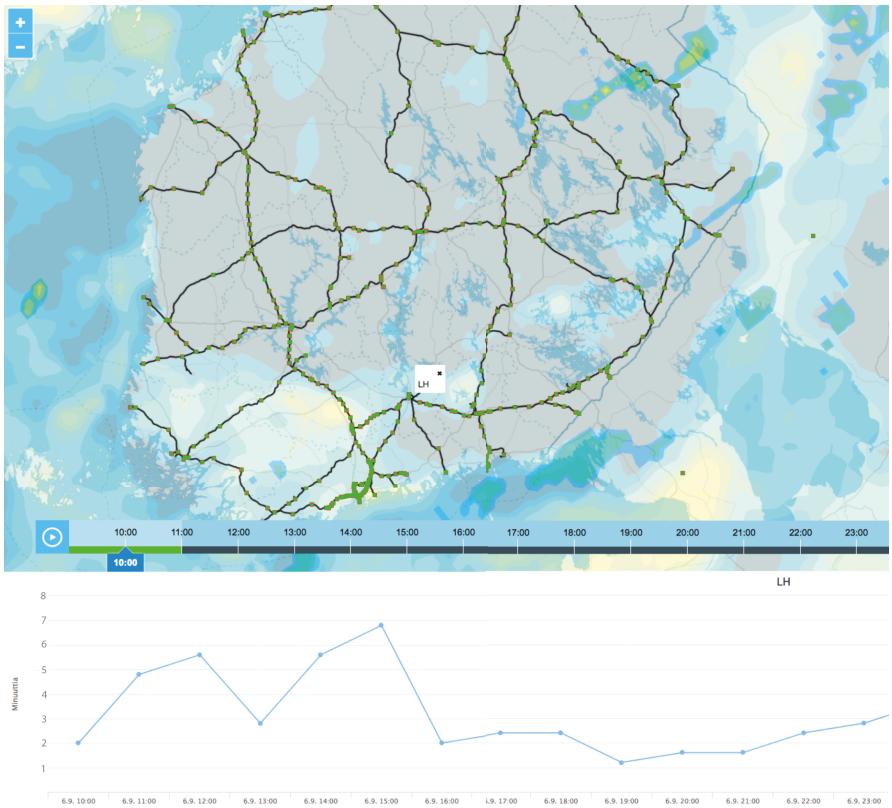

Figure 5.3. UI to explore train delay predictions. The map contains a rail network with colour-coded train stations. The green colour indicates that no severe disruptions are anticipated. The graph under the map shows the predicted amount of delay as a function of time at the selected station. The map in the figure also contains wind gust (light blue areas) and precipitation (green areas) overlays.

6. Summary of Results and Discussion

This thesis studied how to employ machine learning to create serviceable weather impact predictions. In particular, the study scrutinised the power outages caused by highly local convective and large-scale extratropical storms and the weather-inflicted train delays.

A novel, object-oriented, method to predict the power outages caused by convective storms were created in Publication I and Publication II and presented in Section 4.2.1. The method incorporates identifying the storm objects from weather radar data, tracking the objects, predicting their movement 6 hours ahead, and classifying them based on their damage potential. The objects are categorised into 4 classes employing RFC, described in Section 3.5.4, based on the number of presumably affected transformers.

As described in Section 4.2.2, the power outage prediction application was then modified for synoptic-scale extratropical storms in Publication III. Similar to the case of convective storms, the application involves identifying, tracking, and classifying the storms based on their damage potential. Nevertheless, the process was modified for different type of data, required to monitor and predict the broad and long-lasting storms. Publication III experimented with several machine learning methods in the classification task. The best evaluation results were obtained with the GP classifier (see Section 3.5.8). The ERA5 reanalysis data, introduced in Section 2.2, were used in the Publication III while operational setting naturally requires operational NWP data.

The weather-inflicted train delays were predicted in Publication IV and presented in Chapter 5. Both binary classifier and regression was created. The main challenge was the insufficient training data as the reasons for the delays have not been recorded consistently in Finland. From the compared machine learning methods, the best classification metrics were obtained using GBC and the best regression results using RFR.

6.1 Key Results

The presented object-oriented method anticipates the power outages caused by convective storms eminently well, and its modified version provides reasonable results for extratropical storms. The gained performance and the analysis are depicted in Chapters 4.2.1 and 4.2.2. Publication I and Publication II provide a strong evidence that classifying storm objects instead of grid points (i.e., [132]), predefined regions (i.e., [86]) or individual point of interests, such as transformers in the power grid (i.e., [131]), is a vindicable approach. The same approach can also be successfully used to predict power outages caused by extratropical storms, while other approaches might function better with geographically broad storms, as discussed in Section 6.2.

Presumably, the method can be applied to any other domain, such as traffic accidents [33, 34], or air traffic disruption [31], where the impacts are quantitative and produced by identifiable weather events. Thus, the availability of the sufficient impact data is the key issue to create valuable impact forecasts.

The performance of the train delay prediction, discussed in Chapter 5, was not as good as in the other applications in this thesis. The gained results provided evidence that the weather-inflicted train delays can be predicted days ahead, but train delay data with detailed and coherently recorded reasons are needed to provide such predictions with a reasonable accuracy. Nevertheless, the results are essential indication that the data should be gathered, and the delay prediction can be developed where or when the necessary data are available.

Comparing the performance with currently used methods is a non-trivial task as the prediction paradigm is significantly different and thus omitted in this work. Nevertheless, the power outage prediction has proven useful and is in operational use in Finnish power grid companies. The train delay prediction was also evaluated and seen prominent by the Finnish rail traffic operators in the early phase of the development. Unfortunately, the latest version has not been experimented with in operations due to the end of the project.

From the machine learning methods point of view, the results agree with existing research. RF, GP, and GBT, discussed in Section 3.5, played the most prominent role in providing the predictions. Especially RF showed solid performance for all tasks. Remarkably, deep learning methods did not reach similar performance to other methods, although experimented in Publication I, Publication II, and Publication III. All articles analysed the importance of each feature used in the predictions employing either the decrease in the node impurity in the decision trees or the permutation analysis [90]. Although methods employed highly dependent parameters such as minimum and mean temperature, virtually all features contributed

to the predictions.

6.2 Discussion and Future Directions

The main concern in the results is that despite the target in Chapters 4.2.2 and 5 is to provide a days-ahead forecast, the development and evaluation were conducted with ERA5 reanalysis and observations (described in Chapter 2). In an operational setting, the NWP is, however, required. The final error of the prediction would then consists of two, presumably independent, components: an NWP error and an impact prediction error.

A number of issues justify the use of selected data. Employing the weather data that can be considered as the ground truth in this context enables focusing on the error of the impact prediction model itself. If the NWP were used, the ingredients of errors would become too complicated or impossible to analyse. The second justification relates to the available training data. Training an impact model requires a consistent time series covering several years since a sufficient number of significant weather events is required. Due to the changes in NWP models, no consistent NWP time series exist. Especially local area model Harmonie, used in Nordic countries, has gone through significant changes during the previous years [171]. The third reason is practical: using ERA5 and observations makes the effort affordable. Gathering all necessary data requires much work and rather a long period of computational time. Acquiring the corresponding NWP time series would require even more work and at least months of computational time as the data is stored in the deep, slow archive.

Another consideration relating to the evaluation lies in the autocorrelation issue, discussed in Chapter 3. In the context of the convective storm-related power outages (see Section 4.2.1), the test set was selected randomly. Both random and continuous time series strategies were compared when predicting the extratropical storm-inflicted power outages (see Section 4.2.2) and the train delays (see Section 5.2). No significant difference between the evaluation strategies was found in the task of power outage prediction. The train delay models provided better results for the randomly selected test set. Nevertheless, as noted in Section 5.2, an extraordinarily challenging period was selected as the test set. Presumably, both autocorrelation and easier random test set contribute to the difference in the evaluation scores.

While providing an intuitive and well-known view to the model performance, the used metrics contain also a significant deficiency, as they fail to measure a scale of the error. In power outage prediction, for example, it is significantly larger error to predict class 0 than class 2 in case of true class 3. All used metrics treat, however, all misclassifications as the same. Possible resolution to this problem might be for example an extension of

AUC for more than two ordinal classes [172].

As already alleged above, the object-orientation employed in Publication I and Publication II is evidently a good choice for predicting the impacts of convective storms. The approach has several advantages compared to the previous solutions employing ground observations, NWP data, or weather radar data as a grid. Providing only relevant parts of the data to the machine learning methods is beneficial, especially for methods, such as RFC or SVC, that are not doing feature learning. Moreover, processing objects instead of the whole NWP or radar data grid decreases the computational complexity of the training, still improving the spatial accuracy of the processed data compared to the NWP or ground observation stations. The object tracking provides information about storm attributes such as age, size, and movement. The objects are also convenient in designing user interfaces and related applications, such as alerting.

While the same advantages are valid for the geospatially wide extratropical storms-related power outages, discussed in Section 4.2.2 and Publication III, preprocessing the data into as objects is not necessarily the best approach. The aggregated attributes of the storm objects may decrease the classification performance, especially when predictive features fluctuate significantly under object coverage. Various machine learning methods, i.e., neural networks, could exploit the local features and utilise 3-dimensional data.

Comparing deep neural networks in the extratropical storms-related power outage prediction and the train delay prediction with methods proposed in Publication III and Publication IV is one of the most compelling follow-up research for this thesis. The deficient label data seemed to be an issue in the train delay prediction model training in Chapter 5. Retraining the delay prediction when or where long enough coherent delay time series with reason codes is available would be attractive.

One presumable shortcoming of the extratropical storm-related power outage and the train delay predictions is the absence of 3-dimensional data. Especially geopotential height at selected pressure levels might improve the quality of the predictions [173]. The upper-level data in the extratropical storms-related power outage was omitted to optimise the effort within a limited time in acquiring the data and developing and optimising the model. Adding parameters from various pressure levels would be an obvious future improvement of this work.

The rationale to use only ground observations in the train delay prediction is, nevertheless, an assumption that available historical upper-air data (such as ERA5) is too different to train the model as the prediction is conducted based on the local area model Harmonie. We argue the geospatial accuracy is more important than the availability of 3-dimensional data. Comparing the different data sources would naturally be an interesting avenue for future exploration.

As noted in Publication IV, we tried to add an aggregation of the weather parameters over a 72 hours time window as predictive features in the train delay prediction without an improvement. Using lagged data might, however, improve the prediction accuracy. Using crowdsourced weather observations might also improve the predictions as they are distributed geospatially much denser than the traditional ground observations. However, at the time of the project, the crowdsourced observations were not available for the authors.

Last, the uncertainty of the predictions was not exposed to the end-users. Especially expert users and duty forecasters would, however, benefit from the uncertainty information, as discussed in Chapter 1. Providing the information is a significant further development of the applications. However, the presentation needs to be designed carefully not to confuse non-expert users [17] and to take both weather forecast and impact prediction uncertainty into account. One possible way to achieve that would be to create several scenarios based on the different EPS weather forecast data members, discussed in Section 2.2.

References

[1] World Meteorological Organization (WMO), "WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services," WMO-No. 1150. 2015.

[2] G. Anderson, H. K. Kootval, D. W. Kull, J. Clements, Stratus Consulting, G. Fleming, M. Éireann, T. Frei, J. K Lazo, D. Letson, et al., "Valuing weather and climate: Economic assessment of meteorological and hydrological services," Tech. Rep., The World Bank, 2015.

[3] M. Mizutori and D. Guha-Sapir, "The human cost of disasters - an overview of the last 20 years 2000-2019," Tech. Rep., Centre for Research on the Epidemiology of Disasters (CRED), 2020.

[4] G. Nabuurs, M. Lindner, P. J Verkerk, K. Gunia, P. Deda, R. Michalak, and G. Grassi, "First signs of carbon sink saturation in european forest biomass," *Nature Climate Change*, vol. 3, no. 9, pp. 792–796, 2013.

[5] L. Chapman, "Weather and climate risks to road transport," *Infrastructure Asset Management*, vol. 2, no. 2, pp. 58–68, 2015.

[6] E. Niemelä, "KESKEYTYSTILASTO 2017 (i)," Tech. Rep. 2018-06-14 11:51:52.916, Energiateollisuus Ry, Eteläranta 10, 00130 Helsinki, Finland, 2018.

[7] T. Burr, S. Merrifield, D. Duffy, J. Griffiths, S. Wright, and G. Barker, "Reducing passenger rail delays by better management of incidents," *National Audit Office for the Office of Rail Regulation*, 2008.

[8] H. Gregow, A. Laaksonen, and M. E. Alper, "Increasing large scale windstorm damage in Western, Central and Northern European forests, 1951-2010," *Scientific Reports*, vol. 7, apr 2017.

[9] K. Csilléry, G. Kunstler, B. Courbaud, D. Allard, P. Lassegues, K. Haslinger, and B. Gardiner, "Coupled effects of wind-storms and drought on tree mortality across 115 forest stands from the Western Alps and the Jura mountains," *Global change biology*, vol. 23, 06 2017.

[10] R. J. Haarsma, W. Hazeleger, C. Severijns, H. De Vries, A. Sterl, R. Bintanja, G. J. Van Oldenborgh, and H. W. Van Den Brink, "More hurricanes to hit western Europe due to global warming," *Geophysical Research Letters*, vol. 40, no. 9, pp. 1783–1788, 2013.

[11] B. Gardiner, K. Blennow, J.M. Carnus, P. Fleischner, F. Ingemarson, G. Landmann, M. Lindner, M. Marzano, B. Nicoll, C. Orazio, J.L. Peyron, M.P. Reviron, M. Schelhaas, A. Schuck, M. Spielmann, and T. Usbeck, *Destructive storms in European Forests: Past and Forthcoming Impacts. Final report*

to European Commission - DG Environment,, European Forest Institute, 2010.

[12] U. Ulbrich, G. C. Leckebusch, and J. G. Pinto, "Extra-tropical cyclones in the present and future climate: A review," in *Theoretical and Applied Climatology*. jan 2009, vol. 96, pp. 117–131, Springer Wien.

[13] M. G. Donat, G. C. Leckebusch, S. Wild, and U. Ulbrich, "Future changes in European winter storm losses and extreme wind speeds inferred from GCM and RCM multi-model simulations," *Natural Hazards and Earth System Sciences*, vol. 11, no. 5, pp. 1351–1370, 2011.

[14] J. G. Pinto, N. Bellenbaum, M. K. Karremann, and P. M. Della-Marta, "Serial clustering of extratropical cyclones over the North Atlantic and Europe under recent and future climate conditions," *Journal of Geophysical Research: Atmospheres*, vol. 118, no. 22, pp. 12,476–12,485, 2013.

[15] J. Barredo, "No upward trend in normalised windstorm losses in europe: 1970-2008," *Natural Hazards and Earth System Sciences*, vol. 10, 01 2010.

[16] D. Rogers and V. Tsirkunov, "Costs and benefits of early warning systems," *Global Assessment Report. United Nations International Strategy for Disaster Reduction*, 2011.

[17] Thomas Kox, Harald Kempf, Catharina Lüder, Renate Hagedorn, and Lars Gerhold, "Towards user-orientated weather warnings," *International journal of disaster risk reduction*, vol. 30, pp. 74–80, 2018.

[18] A. Mosavi, P. Ozturk, and K. Chau, "Flood prediction using machine learning models: Literature review," *Water*, vol. 10, no. 11, pp. 1536, 2018.

[19] P. Barmpoutis, P. Papaioannou, K. Dimitropoulos, and N. Grammalidis, "A review on early forest fire detection systems using optical remote sensing," *Sensors*, vol. 20, no. 22, pp. 6442, 2020.

[20] E. Hart, K. Sim, K. Kamimura, C. Meredieu, D. Guyon, and B. Gardiner, "Use of machine learning techniques to model wind damage to forests," *Agricultural and forest meteorology*, vol. 265, pp. 16–29, 2019.

[21] R. Lagerquist, M. D Flannigan, X. Wang, and G. A. Marshall, "Automated prediction of extreme fire weather from synoptic patterns in northern Alberta, Canada," *Canadian Journal of Forest Research*, vol. 47, no. 9, pp. 1175–1183, 2017.

[22] R. Chen, W. Zhang, and X. Wang, "Machine learning in tropical cyclone forecast modeling: A review," *Atmosphere*, vol. 11, no. 7, pp. 676, 2020.

[23] A. Wimmers, C. Velden, and J. H. Cossuth, "Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery," *Monthly Weather Review*, vol. 147, no. 6, pp. 2261–2282, 2019.

[24] J. L. Cintineo, M. J. Pavolonis, J. M. Sieglaff, and D. T. Lindsey, "An empirical model for assessing the severe weather potential of developing convection," *Weather and Forecasting*, vol. 29, no. 3, pp. 639–653, 2014.

[25] P. J. Rossi, *Object-Oriented Analysis and Nowcasting of Convective Storms in Finland; Ukkosrajuilmojen oliopohjainen analyysi ja lähihetkiennustaminen Suomessa*, Aalto University publication series DOCTORAL DISSERTATIONS; 164/2015. Aalto University; Aalto-yliopisto, 2015.

[26] G. R. Herman and R. S. Schumacher, "Money doesn't grow on trees, but forecasts do: Forecasting extreme precipitation with random forests," *Monthly Weather Review*, vol. 146, no. 5, pp. 1571–1600, 2018.

[27] D. J. Gagne II, S. E. Haupt, D. W Nychka, and G. Thompson, "Interpretable deep learning for spatial analysis of severe hailstorms," *Monthly Weather Review*, vol. 147, no. 8, pp. 2827–2845, 2019.

[28] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins, et al., "Application of deep convolutional neural networks for detecting extreme weather in climate datasets," *arXiv preprint arXiv:1605.01156*, 2016.

[29] A. Chattopadhyay, P. Hassanzadeh, and S. Pasha, "Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data," *Scientific Reports*, vol. 10, no. 1, pp. 1–13, 2020.

[30] E. Racah, C. Beckham, T. Maharaj, S. Ebrahimi Kahou, M. Prabhat, and C. Pal, "Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events," *Advances in Neural Information Processing Systems*, vol. 30, pp. 3402–3413, 2017.

[31] L. Carvalho, A. Sternberg, L. M. Gonçalves, A. B. Cruz, A. J. Soares, D. Brandão, D. Carvalho, and E. Ogasawara, "On the relevance of data science for flight delay research: a systematic review," *Transport Reviews*, vol. 0, no. 0, pp. 1–30, 2020.

[32] F. Yanovsky, Y. Ostrovsky, and V. Marchuk, "Algorithms for object recognition by weather radar," in *2008 International Radar Symposium*. IEEE, 2008, pp. 1–4.

[33] C. Gutierrez-Osorio and C. Pedraza, "Modern data sources and techniques for analysis and forecast of road accidents: A review," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 4, pp. 432 – 446, 2020.

[34] P. B. Silva, M. Andrade, and S. Ferreira, "Machine learning applied to road safety modeling: A systematic literature review," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 6, pp. 775 – 790, 2020.

[35] Y. Liu and J. Zhong, "Risk assessment of power systems under extreme weather conditions — a review," in *2017 IEEE Manchester PowerTech*, 2017, pp. 1–6.

[36] A. Mellit, A. Massi Pavan, E. Ogliari, S. Leva, and V. Lughi, "Advanced methods for photovoltaic output power forecasting: A review," *Applied Sciences*, vol. 10, no. 2, 2020.

[37] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in *2011 IEEE international conference on smart grid communications (SmartGridComm)*. IEEE, 2011, pp. 528–533.

[38] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian, "A critical review of wind power forecasting methods—past, present and future," *Energies*, vol. 13, no. 15, 2020.

[39] KM J. Rahman, M. M. Munnee, and S. Khan, "Largest blackouts around the world: Trends and data analyses," in *2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*. IEEE, 2016, pp. 155–159.

[40] J. Ludvigsen and R. Klæboe, "Extreme weather impacts on freight railways in Europe," *Natural hazards*, vol. 70, no. 1, pp. 767–787, 2014.

[41] U. Juntti, "Impact of climate on railway operation: a Swedish case study," in *International Heavy Haul Association Conference: 19/06/2011-22/06/2011*. International Heavy Haul Association, 2012.

[42] R. Molarius, V. Könönen, P. Leviäkangas, J. Rönty, A-M. Hietajärvi, K Oiva, et al., "The extreme weather risk indicators (EWRI) for the European transport system," *Natural hazards*, vol. 72, no. 1, pp. 189–210, 2014.

[43] E. Nagy and C. Csiszár, "Analysis of delay causes in railway passenger transportation," *Periodica Polytechnica Transportation Engineering*, vol. 43, no. 2, pp. 73–80, 2015.

[44] Y. Xia, J.N. Van Ommeren, P. Rietveld, and W. Verhagen, "Railway infrastructure disturbances and train operator performance: The role of weather," *Transportation research part D: transport and environment*, vol. 18, pp. 97–102, 2013.

[45] X. Ling, Y. Peng, S. Sun, P. Li, and P. Wang, "Uncovering correlation between train delay and train exposure to bad weather," *Physica A: Statistical Mechanics and its Applications*, vol. 512, pp. 1152–1159, 2018.

[46] L. Oneto, E. Fumeo, G. Clerico, R. Canepa, F. Papa, C. Dambra, N. Mazzino, and D. Anguita, "Advanced analytics for train delay prediction systems by including exogenous weather data," in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2016, pp. 458–467.

[47] P. Huang, C. Wen, L. Fu, L. Lessan, C. Jiang, Q. Peng, and X. Xu, "Modeling train operation as sequences: A study of delay prediction with operation and weather data," *Transportation Research Part E: Logistics and Transportation Review*, vol. 141, pp. 102022, 2020.

[48] S. Barrell, L. P. Riishojgaard, and J. Dibbern, "The global observing system," World Meteorological Organization (WMO) Bulletin nº : Vol 62 (1). 2013.

[49] R. E. López, "The lognormal distribution and cumulus cloud populations," *Monthly Weather Review*, vol. 105, no. 7, pp. 865–872, 1977.

[50] G. B. Foote and C. G. Mohr, "Results of a randomized hail suppression experiment in northeast Colorado. Part VI: Post hoc stratification by storm intensity and type," *Journal of Applied Meteorology and Climatology*, vol. 18, no. 12, pp. 1589–1600, 1979.

[51] J. Rauhala, H. E. Brooks, and D. M. Schultz, "Tornado climatology of Finland," *Monthly weather review*, vol. 140, no. 5, pp. 1446–1456, 2012.

[52] E. Saltikoff, J. Tuovinen, J. Kotro, T. Kuitunen, and H. Hohti, "A climatological comparison of radar and ground observations of hail in Finland," *Journal of applied meteorology and climatology*, vol. 49, no. 1, pp. 101–114, 2010.

[53] V. N. Bringi and V. Chandrasekar, *Polarimetric Doppler weather radar: principles and applications*, Cambridge university press, 2001.

[54] R. K. Crane, "Automatic cell detection and tracking," *IEEE Transactions on geoscience electronics*, vol. 17, no. 4, pp. 250–262, 1979.

[55] M. Dixon and G. Wiener, "TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting—A Radar-based Methodology," *Journal of Atmospheric and Oceanic Technology*, vol. 10, no. 6, pp. 785–797, 1993.

[56] J. T. Johnson, P. L. MacKeen, A. Witt, E. D. W. Mitchell, G. J. Stumpf, M. D. Eilts, and K. W. Thomas, "The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm," *Weather and Forecasting*, vol. 13, no. 2, pp. 263 – 276, 01 Jun. 1998.

[57] A. M. Hering, U. Germann, M. Boscacci, and S. Sénési, "Operational nowcasting of thunderstorms in the Alps during MAP D-PHASE," in *Proceedings of 5th European Conference on Radar in Meteorology and Hydrology (ERAD)*. Copernicus Göttingen, Germany, 2008, vol. 30, pp. 1–5.

[58] H. Kyznarová and P. Novák, "Celltrack—convective cell tracking algorithm and its use for deriving life cycle characteristics," *Atmospheric research*, vol. 93, no. 1-3, pp. 317–327, 2009.

[59] L. Han, S. Fu, L. Zhao, Y. Zheng, H. Wang, and Y. Lin, "3D convective storm identification, tracking, and forecasting—An enhanced TITAN algorithm," *Journal of Atmospheric and Oceanic Technology*, vol. 26, no. 4, pp. 719–732, 2009.

[60] K. Kober and A. Tafferner, "Tracking and nowcasting of convective cells using remote sensing data from radar and satellite," *Meteorologische Zeitschrift*, vol. 1, pp. 75–84, 2009.

[61] S. Shimizu and H. Uyeda, "Algorithm for the identification and tracking of convective cells based on constant and adaptive threshold methods using a new cell-merging and-splitting scheme," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 90, no. 6, pp. 869–889, 2012.

[62] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: algorithms and complexity*, Courier Corporation, 1998.

[63] L. Li, W. Schmid, and J. Joss, "Nowcasting of motion and growth of precipitation with radar over a complex orography," *Journal of applied meteorology and climatology*, vol. 34, no. 6, pp. 1286–1300, 1995.

[64] B. K. P. Horn and G. B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[65] C. Pierce, A. Seed, S. Ballard, D. Simonin, and Z. Li, "Nowcasting," in *Doppler Radar Observations*, Joan Bech and Jorge Luis Chau, Eds., chapter 4. IntechOpen, Rijeka, 2012.

[66] N. E. Bowler, C. E. Pierce, and A. W. Seed, "STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP," *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, vol. 132, no. 620, pp. 2127–2155, 2006.

[67] L. Foresti, I. V. Sideris, D. Nerini, L. Beusch, and U. Germann, "Using a 10-year radar archive for nowcasting precipitation growth and decay: A probabilistic machine learning approach," *Weather and Forecasting*, vol. 34, no. 5, pp. 1547 – 1569, 01 Oct. 2019.

[68] S. Samsi, C. J. Mattioli, and M. S. Veillette, "Distributed deep learning for precipitation nowcasting," in *2019 IEEE High Performance Extreme Computing Conference (HPEC)*, 2019, pp. 1–7.

[69] G Ayzel, M Heistermann, A Sorokin, O Nikitin, and O Lukyanova, "All convolutional neural networks for radar-based precipitation nowcasting," *Procedia Computer Science*, vol. 150, pp. 186–192, 2019.

[70] R. Prudden, S. Adams, D. Kangin, N. Robinson, S. Ravuri, S. Mohamed, and A. Arribas, "A review of radar-based nowcasting of precipitation and applicable machine learning techniques," *arXiv preprint arXiv:2005.04988*, 2020.

[71] M. Peura, "Computer vision methods for anomaly removal," in *Proc. ERAD*, 2002, vol. 2002, pp. 312–317.

[72] S. M. Govorushko, *Natural processes and human impacts: Interactions between humanity and the environment*, Springer Science & Business Media, 2011.

[73] I. M. Navon, *Data Assimilation for Numerical Weather Prediction: A Review*, pp. 21–65, Springer, Berlin, Heidelberg, 2009.

[74] P. Lynch, *The emergence of numerical weather prediction: Richardson's dream*, Cambridge University Press, 2006.

[75] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J. Thépaut, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, 2020.

[76] A. Jung, "Machine learning. the basics.," https://github.com/alexjungaalto/MachineLearningTheBasics, 2021.

[77] S. Karthick, D. Malathi, J. S. Sudarsan, and S. Nithiyanantham, "Performance, evaluation and prediction of weather and cyclone categorization using various algorithms," *Modeling Earth Systems and Environment*, pp. 1–9, 2020.

[78] J. P. Kossin and M. Sitkowski, "An objective model for identifying secondary eyewall formation in hurricanes," *Monthly Weather Review*, vol. 137, no. 3, pp. 876–892, 2009.

[79] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization.," *Journal of machine learning research*, vol. 13, no. 2, 2012.

[80] D. R. Roberts, V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F. Dormann, "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," *Ecography*, vol. 40, no. 8, pp. 913–929, 2017.

[81] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220–239, 2017.

[82] S. Wang and X. Yao, "Diversity analysis on imbalanced data sets by using ensemble models," in *2009 IEEE symposium on computational intelligence and data mining*. IEEE, 2009, pp. 324–331.

[83] T. F. Chan, G. H. Golub, and R. J. LeVeque, "Updating formulae and a pairwise algorithm for variances computing sample," in *COMPSTAT*. Springer Science & Business Media, 1979, p. 30.

[84] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society: Series A (General)*, vol. 135, no. 3, pp. 370–384, 1972.

[85] Z. Li, A. Singhee, H. Wang, A. Raman, S. Siegel, F. Heng, R. Mueller, and G. Labut, "Spatio-temporal forecasting of weather-driven damage in a distribution system," in *2015 IEEE Power & Energy Society General Meeting*. IEEE, 2015, pp. 1–5.

[86] A. Singhee and H. Wang, "Probabilistic forecasts of service outage counts from severe weather in a distribution grid," in *2017 IEEE Power & Energy Society General Meeting*. IEEE, 2017, pp. 1–5.

[87] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, p. 309, Springer series in statistics New York, 2001.

[88] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*, CRC press, 1984.

[89] D. W. Wanik, E. N. Anagnostou, B. M. Hartman, M. E. B. Frediani, and M. Astitha, "Storm outage modeling for an electric distribution network in northeastern USA," *Natural Hazards*, vol. 79, no. 2, pp. 1359–1384, 2015.

[90] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[91] A. McGovern, K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, "Using artificial intelligence to improve real-time decision-making for high-impact weather," *Bulletin of the American Meteorological Society*, vol. 98, no. 10, pp. 2073–2090, 2017.

[92] D. Cerrai, D. W. Wanik, M. A. E. Bhuiyan, X. Zhang, J. Yang, M. E. B. Frediani, and E. N. Anagnostou, "Predicting storm outages through new representations of weather and vegetation," *IEEE Access*, vol. 7, pp. 29639–29654, 2019.

[93] S. Shield, "Predictive modeling of thunderstorm-related power outages," M.S. thesis, The Ohio State University, 2018.

[94] J. H. Friedman, "Stochastic gradient boosting," *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[95] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *Journal-Japanese Society For Artificial Intelligence*, vol. 14, no. 771-780, pp. 1612, 1999.

[96] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[97] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, pp. 164–223, MIT Press, 2016.

[98] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*, pp. 211–230, Cambridge university press, 2004.

[99] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, Jan. 2006.

[100] L. Bottou and C. Lin, "Support vector machine solvers," *Large scale kernel machines*, vol. 3, no. 1, pp. 301–320, 2007.

[101] IBM, "Weather Company Outage Prediction," Cited 14 Aug 2021. Available at: https://www.ibm.com/fi-en/products/outage-prediction.

[102] W. Stanley, "Alternating-current development in America," *Journal of the Franklin Institute*, vol. 173, no. 6, pp. 561–580, 1912.

[103] T. H. Haines, "V—Hurricane experiences of power utilities in Boston Area," *Electrical Engineering*, vol. 58, no. 3, pp. 109–110, 1939.

[104] R. Billinton, *Power system reliability evaluation*, Taylor & Francis, 1970.

[105] R. Billinton and L. Wenyuan, "A novel method for incorporating weather effects in composite system adequacy evaluation," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 1154–1160, 1991.

[106] R. Billinton and G. Singh, "Application of adverse and extreme adverse weather: modelling in transmission and distribution system reliability evaluation," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 153, no. 1, pp. 115–120, 2006.

[107] R Billinton and L Cheng, "Incorporation of weather effects in transmission system models for composite system adequacy evaluation," in *IEE Proceedings C (Generation, Transmission and Distribution)*. IET, 1986, vol. 133, pp. 319–327.

[108] Roy Billinton and Kenneth E Bollinger, "Transmission system reliability evaluation using markov processes," *IEEE Transactions on power apparatus and systems*, , no. 2, pp. 538–547, 1968.

[109] K. A. Clements, B. P. Lam, D. J. Lawrence, T. A. Mikolinnas, N. D. Reppen, R. J. Ringlee, and B. F. Wollenberg, "Transmission system reliability methods-1. mathematical models, computing methods, and results.," *Electric Power Research Institute,(Report) EPRI EL*, vol. 1, pp. var–paging, 1982.

[110] N. Balijepalli, S. S. Venkata, C. W. Richter, R. D. Christie, and V. J. Longo, "Distribution system reliability assessment due to lightning storms," *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp. 2153–2159, 2005.

[111] R. H. Stillman, "Probabilistic derivation of overstress for overhead distribution in-line structures," *IEEE Transactions on Reliability*, vol. 43, no. 3, pp. 366–374, 1994.

[112] W. Li, J. Zhou, K. Xie, and X. Xiong, "Power system risk assessment using a hybrid method of fuzzy set and Monte Carlo simulation," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 336–343, 2008.

[113] J. Wang, X. Xiong, Z. Li, W. Wang, and J. Zhu, "Wind forecast-based probabilistic early warning method of wind swing discharge for OHTLs," *IEEE Transactions on Power Delivery*, vol. 31, no. 5, pp. 2169–2178, 2016.

[114] Q. Zhou, J. Zhang, Z. Yang, and Y. Yang, "Distribution network outage pre-warning analytics under extreme weather conditions," in *2014 China International Conference on Electricity Distribution (CICED)*, 2014, pp. 1190–1194.

[115] E. Brostrom, J. Ahlberg, and L. Soder, "Modelling of ice storms and their impact applied to a part of the Swedish transmission network," in *2007 IEEE Lausanne Power Tech*. IEEE, 2007, pp. 1593–1598.

[116] L. Makkonen, "Modeling power line icing in freezing precipitation," *Atmospheric research*, vol. 46, no. 1-2, pp. 131–142, 1998.

[117] H. Yang, C. Y. Chung, J. Zhao, and Z. Dong, "A probability model of ice storm damages to transmission facilities," *IEEE Transactions on power delivery*, vol. 28, no. 2, pp. 557–565, 2013.

[118] K. Savadjiev and M. Farzaneh, "Modeling of icing and ice shedding on overhead power lines based on statistical analysis of meteorological data," *IEEE Transactions on power delivery*, vol. 19, no. 2, pp. 715–721, 2004.

[119] Y. Zhou, A. Pahwa, and S. S. Yang, "Modeling weather-related failures of overhead distribution lines," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1683–1690, 2006.

[120] R. Eskandarpour and A. Khodaei, "Machine learning based power grid outage prediction in response to extreme events," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3315–3316, 2017.

[121] S. D. Guikema, R. Nateghi, S. M. Quiring, A. Staid, A. C. Reilly, and M. Gao, "Predicting Hurricane Power Outages to Support Storm Response Planning," *IEEE Access*, vol. 2, pp. 1364–1373, 2014.

[122] G. Wang, T. Xu, T. Tang, T. Yuan, and H. Wang, "A Bayesian network model for prediction of weather-related failures in railway turnout systems," *Expert Systems with Applications*, vol. 69, pp. 247–256, 2017.

[123] P. Chen and M. Kezunovic, "Fuzzy logic approach to predictive risk analysis in distribution outage management," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2827–2836, 2016.

[124] Y. Liu, Y. Zhong, and Q. Qin, "Scene Classification Based on Multiscale Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 12, pp. 7109 – 7121, 7 2018.

[125] S. A. B. de Almeida, C. Loureiro, F. P. M. Barbosa, and R. Pestana, "Historical data analysis of lightning and its relation with the Portuguese Transmission system outages," in *2009 IEEE Bucharest PowerTech*. IEEE, 2009, pp. 1–8.

[126] E. Savory, G. A. R. Parke, M. Zeinoddini, N. Toy, and P. Disney, "Modelling of tornado and microburst-induced wind loading and failure of a lattice transmission tower," *Engineering structures*, vol. 23, no. 4, pp. 365–375, 2001.

[127] H. Hangan, E. Savory, A. El Damatty, J. Galsworthy, and C. Miller, "Modeling and prediction of failure of transmission lines due to high intensity winds," in *Structures Congress 2008: Crossing Borders*, 2008, pp. 1–8.

[128] D. Zhu, D. Cheng, R. P. Broadwater, and C. Scirbona, "Storm modeling for prediction of power distribution system outages," *Electric power systems research*, vol. 77, no. 8, pp. 973–979, 2007.

[129] P. Kankanala, A. Pahwa, and S. Das, "Regression models for outages due to wind and lightning on overhead distribution feeders," in *Power and Energy Society General Meeting, 2011 IEEE*. IEEE, 2011, pp. 1–4.

[130] P. Kankanala, A. Pahwa, and S. Das, "Estimation of Overhead Distribution System Outages Caused by Wind and Lightning Using an Artificial Neural Network," in *International Conference on Power System Operation & Planning*, 2012.

[131] P. Kankanala, S. Das, and A. Pahwa, "AdaBoost$^+$: An Ensemble Learning Approach for Estimating Weather-Related Outages in Distribution Systems," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 359–367, 2014.

[132] M. Yue, T. Toto, M. P. Jensen, S. E. Giangrande, and R. Lofaro, "A Bayesian approach-based outage prediction in electric utility systems using radar measurement data," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6149–6159, 2018.

[133] J. He, D. W. Wanik, B. M. Hartman, E. N. Anagnostou, M. Astitha, and M. E. B. Frediani, "Nonparametric tree-based predictive modeling of storm outages on an electric distribution network," *Risk Analysis*, vol. 37, no. 3, pp. 441–458, 2017.

[134] F. Yang, D. W. Wanik, D. Cerrai, M. A. E. Bhuiyan, and E. N. Anagnostou, "Quantifying uncertainty in machine learning-based power outage prediction model training: A tool for sustainable storm restoration," *Sustainability*, vol. 12, no. 4, pp. 1525, 2020.

[135] P. J. Rossi, V. Hasu, K. Halmevaara, A. Mäkelä, J. Koistinen, and H. Pohjola, "Real-time hazard approximation of long-lasting convective storms using emergency data," *Journal of Atmospheric and Oceanic Technology*, vol. 30, no. 3, pp. 538–555, 2013.

[136] J. Sander, M. Ester, H. Kriegel, and X. Xu, "Density-based clustering in spatial databases: The algorithm gdbscan and its applications," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 169–194, 1998.

[137] P. J. Rossi, V Chandrasekar, V. Hasu, and D. Moisseev, "Kalman filtering–based probabilistic nowcasting of object-oriented tracked convective storms," *Journal of Atmospheric and Oceanic Technology*, vol. 32, no. 3, pp. 461–477, 2015.

[138] J. Matthews and J. Trostel, "An improved storm cell identification and tracking (scit) algorithm based on dbscan clustering and jpda tracking methods," in *21st international lightning detection conference*, April 2010.

[139] P. Rossi and A. Mäkelä, "A clustering-based tracking method for convective cell identification and analysis," in *Proc. Fifth European Conf. on Radar Meteorology (ERAD 2008)*, 2008.

[140] E. Kabir, S. D. Guikema, and S. M. Quiring, "Predicting thunderstorm-induced power outages to support utility restoration," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4370–4381, 2019.

[141] S. D. Guikema and S. M. Quiring, "Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data," *Reliability Engineering & System Safety*, vol. 99, pp. 178–182, 2012.

[142] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[143] W. Kron and A. Schuck, "After the floods," Tech. Rep., Münchener Rückversicherungs-Gesellschaft (Munich Reinsurance Company), 2013.

[144] H. Valta, I. Lehtonen, T. Laurila, A. Venäläinen, M. Laapas, and H. Gregow, "Communicating the amount of windstorm induced forest damage by the maximum wind gust speed in Finland," *Advances in Science and Research*, vol. 16, pp. 31–37, 03 2019.

[145] B. Gardiner, A. Schuck, M. Schelhaas, C. Orazio, K. Blennow, and B. Nicoll, "Living with storm damage to forests what science can tell us what science can tell us," *European Forest Institute, EFI, Joensuu, Finland*, 12 2013.

[146] B. Norheim and A. Ruud, "The underestimated demand for public transport?," in *European Transport Conference 2011Association for European Transport (AET) Transportation Research Board*, 2011.

[147] F. A. Shaw, A. Malokin, P. L. Mokhtarian, and G. Circella, "Who doesn't mind waiting? examining the relationships between waiting attitudes and person-and travel-related attributes," *Transportation*, pp. 1–35, 2019.

[148] J. Cordeau, P. Toth, and D. Vigo, "A survey of optimization models for train routing and scheduling," *Transportation science*, vol. 32, no. 4, pp. 380–404, 1998.

[149] T. Dollevoet, F. Corman, A. D'Ariano, and D. Huisman, "An iterative optimization framework for delay management and train scheduling," *Flexible Services and Manufacturing Journal*, vol. 26, no. 4, pp. 490–515, 2014.

[150] G. Zakeri and N. O. E. Olsson, "Investigation of punctuality of local trains-the case of Oslo area," *Transportation Research Procedia*, vol. 27, pp. 373–379, 2017.

[151] Z. Chen and Y. Wang, "Impacts of severe weather events on high-speed rail and aviation delays," *Transportation research part D: transport and environment*, vol. 69, pp. 168–183, 2019.

[152] P. Xu, F. Corman, and Q. Peng, "Analyzing railway disruptions and their impact on delayed traffic in Chinese high-speed railway," *IFAC-PapersOnLine*, vol. 49, no. 3, pp. 84–89, 2016.

[153] E. Diab and A. Shalaby, "Metro transit system resilience: Understanding the impacts of outdoor tracks and weather conditions on metro system interruptions," *International Journal of Sustainable Transportation*, vol. 14, no. 9, pp. 657–670, 2020.

[154] C. Palmqvist, N. Olsson, and L: Hiselius, "Some influencing factors for passenger train punctuality in Sweden," *International Journal of Prognostics and Health Management*, vol. 8, 2017.

[155] W. Brazil, A. White, M. Nogal, B. Caulfield, A. O'Connor, and C. Morton, "Weather and rail delays: Analysis of metropolitan rail in Dublin," *Journal of Transport Geography*, vol. 59, pp. 69–76, 2017.

[156] P. Leviäkangas and S. Michaelides, "Transport system management under extreme weather risks: views to project appraisal, asset value protection and risk-aware system management," *Natural hazards*, vol. 72, no. 1, pp. 263–286, 2014.

[157] N. Bešinović, "Resilience in railway transport systems: a literature review and research agenda," *Transport Reviews*, vol. 40, no. 4, pp. 457–478, 2020.

[158] A. Berger, A. Gebhardt, M. Müller-Hannemann, and M. Ostrowski, "Stochastic delay prediction in large train networks," in *11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2011.

[159] M. Yaghini, M. M. Khoshraftar, and M. Seyedabadi, "Railway passenger train delay prediction via neural network model," *Journal of advanced transportation*, vol. 47, no. 3, pp. 355–368, 2013.

[160] S. Milinković, M. Marković, S. Vesković, M. Ivić, and N. Pavlović, "A fuzzy petri net model to estimate train delays," *Simulation Modelling Practice and Theory*, vol. 33, pp. 144–157, 2013.

[161] P. Kecman and R. M. P. Goverde, "Online data-driven adaptive prediction of train event times," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 465–474, 2015.

[162] R. M. P. Goverde, "A delay propagation algorithm for large-scale railway traffic networks," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 3, pp. 269–287, 2010.

[163] I. A. Hansen, R. M. P. Goverde, and D. J. van der Meer, "Online train delay recognition and running time prediction," in *13th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2010, pp. 1783–1788.

[164] S. Pongnumkul, T. Pechprasarn, N. Kunaseth, and K. Chaipah, "Improving arrival time prediction of Thailand's passenger trains using historical travel times," in *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2014, pp. 307–312.

[165] P. Kecman, *Models for predictive railway traffic management*, Ph.D. thesis, Delft University of Technology, 2014.

[166] L. Kloow, "High-speed train operation in winter climate," *Transrail Publication BVF5*, vol. 2, pp. 2011, 2011.

[167] G. Wang, T. Xu, T. Tang, T. Yuan, and H. Wang, "A bayesian network model for prediction of weather-related failures in railway turnout systems," *Expert systems with applications*, vol. 69, pp. 247–256, 2017.

[168] I. S. Oslakovic, H. ter Maat, A Hartmann, and D. Geert, "Risk assessment of climate change impacts on railway infrastructure," in *Proceedings – EPOC 2013Con*, 2013.

[169] D. Serdar, S. Kaewunruen, M. An, and J. M. Sussman, "Bayesian network-based probability analysis of train derailments caused by various extreme weather patterns on railway turnouts," *Safety Science*, vol. 110, pp. 20 – 30, 2018, Railway safety.

[170] C. Malandri, A. Fonzone, and O. Cats, "Recovery time and propagation effects of passenger transport disruptions," *Physica A: Statistical Mechanics and its Applications*, vol. 505, pp. 7–17, 2018.

[171] L. Bengtsson, U. Andrae, T. Aspelien, Y. Batrak, J. Calvo, W. de Rooy, W. Gleeson, B. Hansen-Sass, M. Homleid, M. Hortal, K. Ivarsson, G. Lenderink, S. Niemelä, K. P. Nielsen, J. Onvlee, L. Rontu, P. Samuelsson, D. S. Muñoz, A. Subias, S. Tijm, V. Toll, X. Yang, and M. Ø. Køltzow, "The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System," *Monthly Weather Review*, vol. 145, no. 5, pp. 1919 – 1935, 2017.

[172] W. Waegeman, B. De Baets, and L. Boullart, "Roc analysis in ordinal regression learning," *Pattern Recognition Letters*, vol. 29, no. 1, pp. 1–9, 2008.

[173] M Hofstätter and G Blöschl, "Vb cyclones synchronized with the arctic-/north atlantic oscillation," *Journal of Geophysical Research: Atmospheres*, vol. 124, no. 6, pp. 3259–3278, 2019.

Aalto-DD 132/2021

ISBN 978-952-64-0524-7 (printed)
ISBN 978-952-64-0525-4 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Computer Science
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

**DOCTORAL
DISSERTATIONS**