Trend Reports
Seminar on Machine Intelligence
Summer Semester 2021

Klaus Diepold, Sven Gronauer, Matthias Kissel, Mohamed Ali Tnani (Eds.)

February 10, 2022

Abstract

The topic of the *Seminar Machine Intelligence* for the summer term 2021 is *Machine Intelligence for the Environment*.

Currently, students, researchers as well as the wider interested public is all excited about the new findings in the field of Machine Intelligence in general and Neural Nets in particular. The expectations are flying high as new algorithms help to separate images from cats and dogs at a level of performance that exceeds humans' capabilities. Moreover, Machine Intelligence also beat the best human players in games such as GO, Chess or Jeopardy.

In spite of all these exciting developments we are facing a major societal challenge in terms of the Climate Crisis. One starts wondering if we can also employ Machine Intelligence/Artificial Intelligence for something more useful than playing games and sorting pictures. Most notably, we are wondering if MI can contribute to battle the Climate Crisis?

This is the question the students participating in the seminar tried to discuss in summer 2021. The students examine the state of the art use of machine learning methods in the context of oceans, solar energy, smart cities, the transportation sector, collective decision making, animals as well as forestry agriculture. Current trends are analyzed and projected into the future, which targets the question whether Machine Intelligence will help us saving our planet.

Foreword

by Klaus Diepold

Dear Reader,

welcome to the Proceedings of the *Seminar on Machine Intelligence* Summer 2021 edition.

These days the academic community spends significant amount of energy discussing innovations in teaching and learning. In particular, the advent of digital tools causes a frenzy about the introduction of digital innovations in teaching. Such digital innovations include video based teaching materials aka MOOCs (massive open online courses), various concepts for 'inverted classrooms', the use of tablets, clickers, online voting and feedback and other computer-based tools as well as using all sorts of digital media and much more.

I am all for experimenting with new formats of teaching, and a reasonable use of digital tools can be helpful and even inspiring at times. Any new way of teaching, be it digitally supported or based on less technically demanding methods and tools shall be evaluated if they actually help students to earn a better education. This last question taken by itself is already a challenge, because we lack reliable methods to measure the success of education and hence we have a hard time to distinguish between *good* and *better*. Looking at examination results is certainly not enough, just as much as student evaluations are seldom more than an assessment of the students' well-being. These methods are prone to all sorts of secondary effects rendering the results close to useless.

One measure of success that I regard as rather reliable and instructive is to what extend I succeed in activating the students in a class. By activation I mean that students are actually going out on their own acquiring facts and knowledge, digesting and actively discussing the material they found among themselves, without me telling them exactly what they are supposed to be doing. This way they create new knowledge and experiences. And that I would declare a successful education.

The book you hold in hands is the result of one of my educational experiments. The students were set up to collect, acquire, digest and produce new knowledge for themselves. This year we ventured to explore future research directions for Artificial Intelligence when applied to the epic struggle of humanity against the self-inflicted effects of the Climate Crisis. I dearly hope that this seminar also serves for the students as a preparation for choosing a topic for their final project before concluding their Master's degree at TUM. The knowledge students gain this way may not be new to the world or to the scientific community, but it is new to the students and it is active in their minds by virtue of the process they went through. Besides the new knowledge they pick up, they also gain experience in the process of collecting and digesting information, being critical and constructive as well as experiencing the power of communication and intellectual exchange with their peers and hence turning information into knowledge.

One aspect that I find instructive to measure the success of this course format is the amount of effort and time students invest in the course voluntarily, without me, the instructor, urging or requiring them to work harder. They just do it, because they feel inspired and because they are curious. Funny enough, this extra engagement on the students side earned me an exhorting message from my Dean of Study, who felt that our students were overly burdened by the course. This exhortation was the result of the students' course evaluation, where students indicated that they've worked many more hours than accounted for the by the assigned credits. However, the students also acknowledged that they loved the course in spite of the long hours. To me, this is a strong indicator that we did something right. I am exuberantly happy about the outcome of the course, which is exactly this book and I am proud of the students who proved very convincingly that they are maturing academically and that they can create interesting research-related output way beyond reproducing the content of lecture notes.

In spite of this somewhat personally felt success, I still had to promise to the Dean of Study that next year we will return to a format with reduced work load. I am not sure if the students can keep up to this promise if I succeed to fire them up to a similar amount, possible with one of my next educational experiments.

Munich, September 2021

Klaus Diepold

Contents

Chapter 1

Introduction

by
Matthias Kissel,
Mohammed Tnani,
Sven Gronauer
and Klaus Diepold

This introduction provides background information about the genesis of this book, why it exists, how it was conceived and how it was finally produced. This account shall serve to also communicate the didactical concept underlying the process that eventually produced the book.

From this account it is also clear that the content of the book represents a snapshot of current thinking about the use of machine intelligence for evaluating, assessing or even mitigating the effects of the Climate Crisis. This is taken with a perspective as seen by students who are just about to enter the world of science. The book is the result of the one-semester *Seminar Machine Intelligence*. The seminar is a course in the curriculum of the Master of Science in Electrical and Computer Engineering, which is offered by the Faculty of Electrical and Computer Engineering of the Technical University of Munich (TUM). The seminar consists mainly of weekly meetings for 2 hours to discuss and work on the subject. The work is organized as group discussions and team work. Students have to read, write, review and present their findings on a weekly basis. To this end, new digital e-learning tools and methods were employed along with discussion styles such as world-cafe or speed-dating discussions and other styles of organized communication. Students do presentations as Pecha Kuchas, a specific presentation format that consists of 20 slides, each shown for 20 seconds. That makes every presentation to last just 6 minutes and 40 seconds. This format facilitate for highly focused and condensed presentation sessions, while also creating a spontaneous and fun atmosphere. It is tough to be boring under such conditions, and if a presenter is boring, it is quickly over. Throughout the seminar participating students should learn fundamental aspects of scientific research along with honing their skills in oral and written communication in a scientific or technical field. The intention for the students in the course is to develop ideas and paths for future research in the field leading towards insights and methods necessary to design and implement intelligent machines in the broader sense of the word.

The seminar in total was structured in three major stages, which we elaborate a bit here.

1st Stage: Individual Reading, Writing, Reviewing, Discussing During the first stage the seminar started out with all students jointly reading a set of fundamental papers or book to set the stage and provide for a shared basis and reference for future discussions. Between subsequent meetings students agreed on a set of chapters to read until the next meeting. Students were also asked to reflect on their reading by writing short essays along some high-level guiding questions. Each of the students' essay had the size of 5000 characters (incl. spaces). The students uploaded their essays before the next meeting using the e-learning platform Moodle. Furthermore, students were randomly assigned essays of their fellow students to read and review. During this stage, the students discussed the content of the book during the weekly meetings, using various forms of discussion, such as world cafe, speed dating discussions, fishbowl discussions, cocktail party discussion. This form of reading, writing, reviewing and discussing generated a shared domain of knowledge to facilitate the later stages in the process. It also conveyed fundamental information about the field of study on intelligent systems as well as some facts on the Climate Crisis. This first stage took about 4 weeks.

2nd Stage: Team - Researching, Presenting, Discussing The second stage started with a workshop where the students tried to identify major fields of science and technology, which were considered essential to push the topic of machine intelligent for handling the Climate Crisis into the future. By the end of the workshop, the student agreed on a list containing the dominating fields and domains. Subsequently, students could assign themselves to one of these items on the list to further study the field in more detail. During the next five weeks, the teams of students researched their chosen field compiling information about the state of the art and the major trends. During the weekly contact hours one student per team delivered a Pecha Kucha presentation, highlighting the group's findings during the past week for all others to understand and participate. The presentations were followed by discussions on open points. The one purpose of the presentations is to disseminate the essence of the information collected over the past week to the fellow students. Another objective is to sharpen the sense of the presenters to think about their target audience and to tailor the amount and the level of detail of the presentations to match the expectation of their target audience.

3rd Stage: Projecting - Writing - Reviewing The book is written by students mainly for students. It does not claim to contain and communicate ultimate truths, but rather tries to project current facts and trends into future directions of research based on a intense investigation of trends and possibilities. The book may prove to be a helpful tool to orient students interested in the study and the development of intelligent systems, AI, machine learning and so on as a basis to narrow down on a topic for their Master thesis projects or even beyond. Not least of it, the book may also display interesting ideas and anticipations, which may be helpful even for more seasoned researchers to communicate with young people and transmit the excitement for science and research on intelligent systems using a language and level that students can digest and appreciate.

We hope, you the reader, will find inspiration in the chapters and the material to further lead the discussion about practical uses of machine intelligence to tackle serious societal challenges. If you have any remarks on this book, our process or the course itself, we would love to hear from you.

Chapter 2

Machine Learning for Oceans

Atamny, Omar
Hacket, Franziska
Li, Wantao;
Lorenz, Julian
Matthies, Stephan
Schuran, Max
Sieber, Konstantin;
Wu, Jiamin
Zhang, Yingyi

Abstract

Our oceans and the life contained within them are largely unexplored, yet they play a critical role in keeping earth habitable for humans and animals alike. Currently, however, large numbers of coral species and other marine life are at risk of extinction as a result of anthropogenic climate change and the pollution of our oceans induced alongside it.

In this report, we investigate problems in marine science and predict prospective future solutions provided by new machine learning algorithms. For this purpose, we analyse different upcoming trends in oceanography. These trends include different forms of data acquisition using autonomous seacraft and remote sensing, for instance in the adoption of unmanned autonomous vehicles and satellite imagery for monitoring events that have a strong climate-impact. Additionally, we detail potential trends for the enhanced prediction of renewable energy sources, which is particularly interesting for the purpose of abandoning fossil fuels. Lastly, we explore how we could obtain better models of our oceans in the future, for instance for mapping- and prediction-purposes. This also includes improved forecasting of extreme weather conditions.

2.1 Introduction

It is said, that over 80% of our oceans are unmapped, unobserved, and unexplored. It is clear that our oceans provide a habitat for billions of lifeforms, over 90% of which have not yet been discovered, according to some estimates [1, 2]. Some of these organisms produce oxygen, some of them provide nutrients to other organisms. This might seem like the beginning of a description of a balanced and stable eco-system, when in fact our oceans are incredibly susceptible to outside influences. Too many parts of the biosphere failing will cause a ripple effect on the rest of the system, and with anthropogenic climate change already negatively affecting the health of our coral reefs, oil spills polluting our waters, and the abundant microplastics in our oceans being absorbed by marine organisms, it is up to us to find solutions to these problems before they become irreversible.

One possible approach to these issues is the use of machine learning. Because of it's broad field of applications, machine learning can be adapted to many scientific studies. For instance, current algorithms can forecast the climate more accurately, which has become more and more difficult in recent years [3]. Further, with the support of state-of-the-art satellites and drones, the yet unmapped areas of the ocean can be explored and unforeseen oceanic catastrophes can be contained, before they damage the environment irreversibly.

In the following sections, potential future trends in machine learning and data science connected to oceanography will be identified and assessed in terms of their potential impact on the environment as well as our efforts for ocean preservation. The report concludes with a driver matrix rating the trends we found in terms of impact and uncertainty, as well as some final remarks.

2.2 Trends

2.2.1 Remote Sensing

For researchers and policy makers alike, information gathering and data acquisition are vital to understanding the world around them. Often, this process of collecting the necessary data is made feasible through the use of sensors taking samples on-site. For the case of oceanography, however, this is not always possible. With oceans covering over 70% of earth's landmass, it is simply not possible to keep an up-to-date view of current events from on-site measurements everywhere at once. This problem can be addressed through the use of *remote sensing*, whereby an object of interest can be measured without making physical contact in the process [4]. This way, data can be acquired without having to be on-site. Traditionally, remote sensing on earth is performed using satellite imagery or aerial vehicles and a wide range of applications exists. For instance, mapping earth's surface, detecting fire and even agricultural planning have all been simplified through the use of remote sensing technologies [5, 6, 7]. One application receiving increasing attention is the use of satellite imagery in monitoring the health state of our oceans. In the following sections, facts about the impact of global warming on marine life as well as some state of the art applications countering these effects will be established. Key drivers for the development of remote sensing in the context of ocean monitoring will be highlighted and subsequently, challenges as well as the impact these trends have on the environment will be discussed.

Facts

- Coral reefs provide habitat for many sea-dwelling species. However, about 50% of all coral reefs have already disappeared [8], and up to 90% of them, including the Great Barrier Reef, are projected to disappear by 2050 if global warming continues at the current rate [9]. Assessing the health state of a reef is traditionally very labour intensive because divers need to collect data on-site, even if the data is later automatically analyzed [10].
- Harmful algae blooms, a process characterized by the quick and widespread growth of harmful algae colonies into large-size algae fields [11], are becoming increasingly frequent. This constitutes a problem due to toxic byproducts produced in the process, causing environmental and economic damages in the process [12].
- In the event of an oil spill, swift action is necessary, because just one litre of oil can potentially contaminate up to a million litres of water [13].

Key Drivers

- In many scientific groups, acquiring data constitutes a large part of the research process. For instance, trained divers often manually capture data under water. In the future, more and more of this work could be diverted to autonomous sub-sea drones and other remote-sensing technologies [14]. Increased use of machine learning methods will only hasten this process and potentially save a considerable amount of resources.
- Recent work has shown that it is possible to detect coral bleaching and assess reef health using satellite imagery in combination with neural-network based approaches [15, 16, 17], which could help in diverting labour force to other important tasks in the future.
- Detecting and understanding the formation of algae blooms is of interest to researchers and policy makers alike. Studies have demonstrated that detecting and assessing them is possible through the use of remote sensing [18]. Machine learning has yet to be widely adopted in this regard, but as with the previous point, it promises a decrease in the labour force needed for this task in the future.
- Swift detection of oil leaks plays a crucial role in preventing irreversible ocean pollution, and researchers are currently testing machine learning methods of detection using satellite imagery in this regard [19, 20, 21].

Challenges

- Clouds can obstruct satellite vision, making it difficult to interpret data reliably.
- Operation time for autonomous drones is limited to due to constraints in battery technology.
- Ensuring time-coherence for satellite data poses a challenge because of the limited time window for which a particular region is in view of the satellite.
- Dataset bias and other signal-processing related issues can potentially lead to unreliable algorithms [21].
- Interference due to oceanic phenomena, like weed beds and algae blooms, makes oil spill detection difficult for some remote sensing techniques [20].

Impact

After talking about the challenges to overcome, this section details the potential impacts of the identified key drivers. With the help of improved satellite imagery, ocean events such as oil spills and coral bleaching events can be detected in the early stages. The broader availability of satellites to researchers makes studying these events easier and could help mitigate further pollution of our oceans. Increased use of machine learning methods will only aid researchers and policy makers in this regard, potentially freeing up a considerable amount of labour force that could be used elsewhere.

2.2.2 Autonomous Seacraft

To gain more knowledge about the complex mechanics of the ocean's ecosystem and physical dynamics, we need to gather data directly from the sea. These so called in-situ measurements are extremely valuable yet difficult to collect. Marine scientists need to take long expeditions on a boat where they take many samples from the ocean. This can be very tedious. As an example, for taking water samples at different depths, manual labour from divers is required. Diving operations require long preparation and can only be performed if the weather allows it. That is why the sample sizes in marine applications are comparatively low. Having only small data sets makes it hard for data scientists to train machine learning algorithms, which in turn could ease the work of marine biologists and physicists. For example, Puissant et al. [22] use a large series of in-situ measurements combined with remote sensing data to train self-organizing maps that predict phytoplankton populations using only satellite data. Because the ocean takes up over 70% of the earth's surface, there is little chance it could be sufficiently sampled only by manual labour. Autonomous seacraft can help to decrease the efforts of taking in-situ samples.

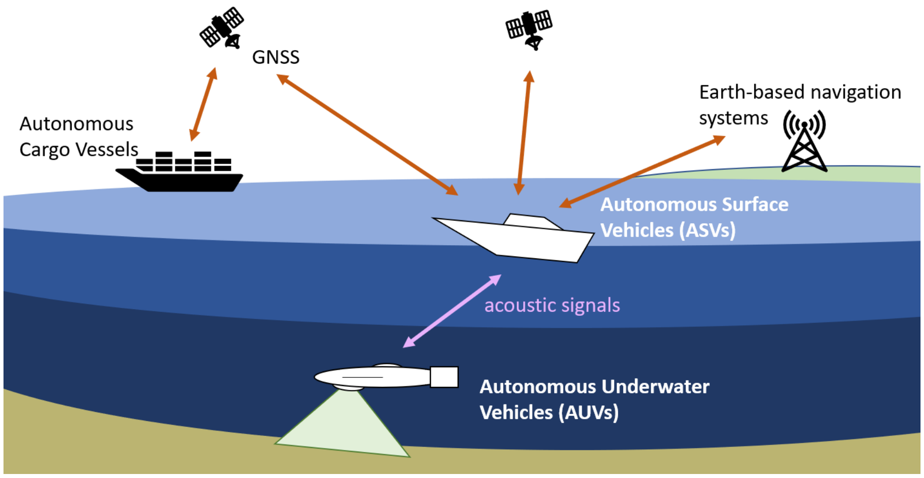

Figure 2.1: Environment of Autonomous Seacrafts

In general, seacraft can be divided in two groups: Remotely Operated Vehicles (ROVs) and Autonomously Operated Vehicles (AOVs). The latter group can be further divided in Autonomous Underwater Vehicles (AUVs), also called Unmanned Underwater Vehicles (UUVs), and Autonomous Surfaces Vehicles (ASVs), also called Autonomous Surface Craft (ASC). In this trend report, we will focus on the autonomous operated vehicles and discuss their challenges and impact on the marine environment and therefore on climate change.

Facts

- There are already several projects for autonomous seacraft as for example FRED - Floating Robot for Eliminating Debris [23]. For these systems, machine learning is included in several parts, for example for sensing the environment with cameras, intelligent energy management and control algorithms.
- For ASVs, GPS is widely available and is sufficient for their navigation requirements.
- According to the US National Oceanic and Atmospheric Administration (NOAA), more than 80% of the ocean's floor remains unmapped, unobserved, and unexplored [2].

Key Drivers

- Ocean floor exploration is one of the main applications for AUVs. Due to the pure surface area of the oceans, it is tedious to cover by manual labour.

The winning technology of the Shell Ocean Discovery, XPRIZE [24], provides an example on how ASVs and AUVs can be used in corporation to take on such tasks.

- Navigation research for vehicles in GPS-denied areas is crucial for AUVs. Advances in swarm navigation research will allow them to determine their position even below the water surface. Reinforcement learning can help at multi-objective path planning, because AUV swarms need to position themselves both optimally for navigation and for exploration at the same time [25].
- Autonomous seacraft can help to identify and report illegal fishing activities [26, 27], for instance by detecting fishing vessels with cameras on the seacraft that use ML-based detection algorithms.
- Current systems for detecting plastic pollution on the oceans can detect plastic on the surface of the oceans. However, most of the plastic is in garbage vortices in the oceans [28]. With AUVs and adapted machine learning algorithms, also the plastic below the surface can be detected.

Challenges

- Especially for AUVs, navigation is still a big challenge as navigation in unconfined environments is still an unsolved problem. Potential solutions could be Simultaneous Localization and Mapping (SLAM) based on bathymetry (i.e. topographical maps based on echosounders) or heterogeneous swarms including a combination of ASVs and AUVs [29].
- Moreover, for ASVs the main source for positioning is GNSS, which is vulnerable for intentional and unintentional RF interference. One example are GPS jammers, which can be actively used to deny access to a position source. Engineers can mitigate that threat by also including other positioning sources such as enhanced Long Range Navigation (eLORAN), Differential GPS (DGPS) or by backing up navigation with inertial measurement units [30].
- Another challenge or even a risk is that with the deeper development of further autonomous seacraft, this will lead to an increase of the production of seacraft as well as to an increase in the usage of seacraft [29]. Depending on the type of drive, more or less greenhouse gas emissions are generated during the production but especially during the usage at the sea. This might lead to Jevon's Paradox as instead of helping the marine environment with autonomous seacraft, autonomous seacraft could even worsen the state of the marine environment.
- On top of the technological challenges, a legal framework for autonomously operating vessels must also be established. It will have to be clarified how autonomous ships are handles under admiralty law [31].

Impact

Autonomous seacraft will improve the way in-situ measurements for ocean environment monitoring are taken. They have the chance to significantly reduce the need for manual labour of scientists and increase the overall sampling rate. This will help to gain a better understanding about the conditions of our oceans and produce large amounts of data which could be used for machine learning applications, for instance to complement remote sensing technologies.

Furthermore, with the help of autonomous seacraft, plastic and oil pollution can be detected and reported easily as well as illegal fishery. In a second step, the autonomous seacraft could also be equipped with a garbage system to directly remove the pollution.

2.2.3 Machine Learning in Marine Renewable Energy Prediction Problems

Climate change's impacts have become abundantly clear. The major reason is that energy generation is reliant on fossil fuels. The most pressing issues in the future years and decades will be the quest for alternative energy sources and the decrease of fossil fuel consumption. Marine renewable energy (MRE) has enormous potential. Marine energy is plentiful and there are large spaces in both marine and coastal environments that can be considered for gathering various kinds of marine energy, such as tide and wave energy, floating solar energy, and osmotic energy. Therefore, innovative technological developments which are relevant to harvesting ocean energy are strongly encouraged [32].

Marine renewable energy generation and its integration into power networks are the areas where machine learning techniques can be applied. Marine natural phenomena such as wind, tides, and waves exhibit the same property: intermittency. Machine learning enables to learn and predict the pattern of intermittency based on large amounts of structured and unstructured data. That is to say, when applying machine learning algorithms to weather reports, including forecasting of sunrise and sunset, Beaufort scale, precipitation, etc., they were able to predict the output over the next 36 hours. However, the integration of natural phenomena with manufactured devices in a harsh environment imposes constraints and implementing these approaches is still a big challenge [33].

Facts

- **Offshore wind energy:**
  Over the previous decade, research into onshore wind has made great achievement. For around ten-fold cost reduction, it becomes more competitive. Wind energy installations in Germany have grown from about 9.5 TWh in 2000 to an approximate 131.9 TWh by the end of 2020 [34]. The cost of offshore wind energy depends on the specific environment and the cost of an offshore project is significantly more than an onshore project [35].

- **Tidal stream and ocean current energy:**
Tidal streams are driven by changes in the tidal head as they move down the coast, and the resources that can be exploited are concentrated in certain locations such as straits, off headlands, in bays, or between islands and landmasses where the coastal geometry contributes to the enhancement of the tidal currents. Ocean currents are formed as a result of oceanic circulation and wind shear. The strongest ocean currents include Florida current and Agulhas current, both currents are capable of reaching speeds of up to $2m \cdot s^{-1}$ [36].

- **Wave energy:**
Engineers have already invented multiple types of the wave energy converter such as salter's duck, pelamis wave energy converter, Manchester Bobber, Archimedes wave swing machine, Wave Dragon, to name some of them. In [36], the author has mentioned that the most recent wave laboratory facilities are Edinburgh's circular wave tank, FloWave, and Shanghai, Plymouth, Cork, Trondheim, and Ghent's rectangular wave basins [36].

Key Drivers

- **Offshore wind energy:**
In comparison to other marine renewable energy sources, research into offshore wind is more mature, owing to the fact that onshore wind energy has already been an important and major industry. On the other hand, offshore winds are much stronger and consistent than onshore wind. Therefore, offshore wind is the best alternative source of onshore wind [37].

- **Tidal stream and ocean current energy:**
Although many variables influence the process of tides, tides are still highly predictable. Tidal and ocean currents often surpass $1m \cdot s^{-1}$ [38].

- **Wave energy:**
Wave energy is the world's largest projected source of ocean energy. The energy density per meter of wave along the beach is between 30 and 40 kW [39].

Challenges

- **Offshore wind energy:**
The research of long-life and cost-effective devices that could survive in the cruel marine environment is still an important challenge. Machine learning can help to reduce costs by catching problems before malfunctions happen and optimize the powerplant schedules based on weather prediction. In [40], the author introduced machine learning methods for wind turbine condition monitoring, which detect malfunctions based on generator temperature and power curve monitoring. As [41] indicated, using inflow wind velocity and directions as predictor variables, a data-driven prediction framework can predict wind farm output by training

machine learning approaches such as the general regression neural network (GRNN), random forest (RF), support vector machine (SVM), gradient boosting regression (GBR), and recurrent neural network (RNN).

- **Tidal stream and ocean current energy:**
  The classical method for the prediction of tidal currents is harmonic analysis, which is deemed as state-of-the-art. However, it has several flaws that could result in forecast errors due to rapid tidal currents and 'noisy' tidal signals. To this end, Dripta Sarkar and his colleagues presented a machine learning approach to the prediction of tidal currents, which is a robust and efficient algorithm by using the Gaussian process. However, this approach is computationally expensive. The research of an alternative algorithm is the current challenge [42].

- **Wave energy:**
  The primary difficulties are optimizing power output and minimizing environmental effects and operational expenses. Using a deep learning technique based on long short-term memory (LSTM), which is introduced in [43], a cost-effective wave energy converter dubbed "Searaser" was created to predict the output accurately.

Figure 2.2: An offshore wind turbine [44]

Impact

- **Offshore wind energy:**
  The cost reduction is expected to advance the spread of offshore wind energy technology, which may become the central pillar of the transition to low-carbon electricity generation.

- **Tidal stream and ocean current energy:**
  It aids in more efficiently harnessing energy from the site of rapid tidal currents.

- **Wave energy:**
  Wave energy is more dependable than wind or solar energy, and so may replace the void created when wind or solar energy fails to function. Thus, they can be seen as a better mix of energy sources, reducing reliance on traditional energy sources such as natural gas, crude oil, and coal.

2.2.4 Extreme Climate Prediction

In recent years, artificial intelligence has been used to help rank climate models, spot cyclones or other large-scale weather events, as well as for the purpose of identifying and predicting climate patterns. However, the question to which extent variable ocean heat is imprinted on the atmosphere to realize its predictive potential over land remains unsolved [45].

El Niño-Southern Oscillation (ENSO) is an inhomogeneous climatic condition, originating in the tropical Pacific Ocean, that causes heavy rainfall, floods and droughts with catastrophic effects in many countries. Long short-term memory (LSTM) models have the ability to capture the long-term time dependence of ENSO, which enables those models to predict ENSO phenomena in multiple steps over a longer period of time. This is done by integrating a complex climate network analysis into a neural network regression task. Figure 2.3 shows the two stages of the mentioned approach. According to the results shown below, the 12 network indicators, which were extracted as predictor variables, have the potential to predict ENSO phonemena over a longer period of time [46].

Sea surface temperature (SST) is used as a prognostic and diagnostic indicator in numerical weather prediction and in global climate modeling [47]. It significantly affects the processes of air-sea interactions and thus forms an important measure of climate changes [48]. Currently, it is possible to make predictions for a given individual location with the help of manual or simple neural networks. Furthermore, there is a growing trend to expand the range of these predictions. As proposed by Xiao et al. [49], artificial neural networks (ANN) using a convolutional long-short-term memory (ConvLSTM) approach can predict SST on a daily time-scale. Selected lengths of past daily SST observations are provided to the neural network and the SST predictions for various points in the future can be derived as outputs. Using this method, ANNs trained on satellite-derived SST data can predict future values up to 7 days in advance and detect extreme weather events up to 5 days in advance [49]. Recent experiments using satellite SST data of the past 36 years in the East China Sea show the superior performance of the proposed model in comparison to the best known algorithms like persistence models, linear support vector regression (SVR) models and two different LSTM models. The results indicate that the model is promising for accurate and convenient short- and medium-term daily SST field forecasting [49].

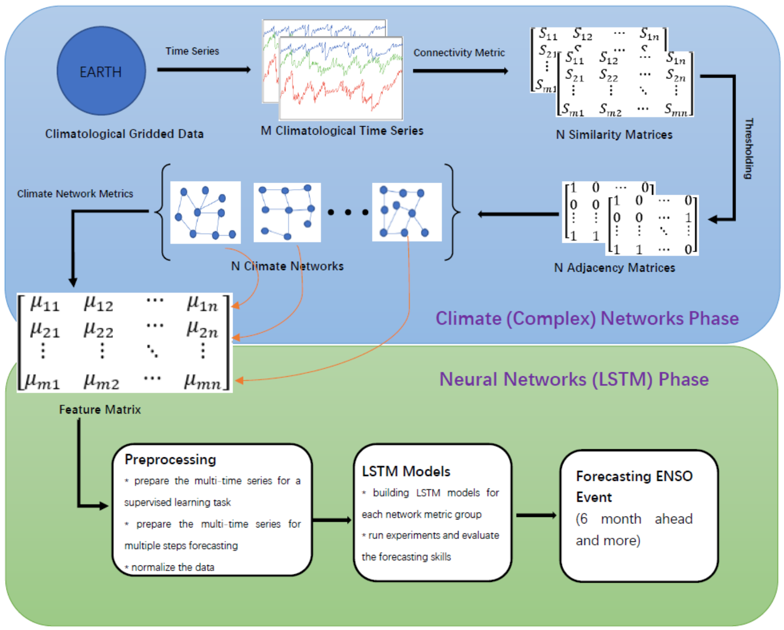

Figure 2.3: The schematic representation of the proposed approach for forecasting ENSO event

Changes in the dissolved oxygen concentration of the ocean have important implications for marine ecosystems and global climate change [50]. The deep learning-based model called *marine deep jointly informed neural network* (M-DJINN), is proposed to predict marine dissolved oxygen concentration (MDOC). According to [50], this new model achieves reduced prediction errors and better convergence in comparison to other models [50]. As the ocean takes on regulatory functions in earth's climate, there is great research potential in applying machine learning to predict typhoons, SST, MDOC and tropical cyclone, to name only a few possible applications.

Facts

- Extreme weather events such as storms, heat waves, floods and droughts increase in frequency and severity. There is a great need to provide early warning of such weather events for the safety of human life and property. However, there is still a lack of accuracy in current extreme weather forecasts.

- Oceanic activity plays a key role in climate change. In addition to long-term effects, the ocean also influences the emergence of extreme weather, which is why great importance is attributed to the study of ocean changes when forecasting extreme climate.

- There are already studies on the use of ocean information to predict extreme weather [51].
- The relationship between the ocean and extreme weather forecasting is not well-studied, which is why current machine learning models mostly do not achieve desirable results.

Key Drivers

- There is a need for extreme weather prediction.
- Research on well-known algorithms and various types of deep learning algorithms.
- Development of various climate prediction models.
- Existing knowledge about the ocean is used to obtain better mathematical models.
- Lessened hardware constraints due to technological advances.

Challenges

- Currently the knowledge about the ocean only scratches the surface. Many more problems have to be solved.
- Due to the vast space of the ocean, any kind of meaningful data acquisition will be very difficult, subsequently affecting the size of captured data.
- As with most scientific fields, interdisciplinary collaboration is essential for achieving satisfactory results.

Impact

Assuming that collaborations between, for instance, oceanography experts, earth science experts and machine learning experts become more common, prediction of extreme weather could improve drastically. If proper use is made of the existing information about our oceans and climate change in general, the arrival of extreme weather events can be predicted much earlier and more accurately. In this way, machine learning can be an important tool in minimizing loss of property and life in the future.

2.2.5 General Modeling and Analysis using Neural Networks

As mentioned in previous sections, oceans cover over 70% of the earth's surface. Naturally, this means that they make up a large portion of the entire ecosystem and are thereby able to exert considerable influence on many events on earth. Owing to the fact that most of our oceans remain unexplored, we neither fully understand how they influence our ecosystem precisely, nor how our actions influence them. Presently, numerical models are dominantly in use when it comes to ocean models or ocean forecasting. These models are computationally expensive and prone to failure, especially when it comes to unstructured, complex natural processes. This is where machine learning models – in this section we focus mostly on Neural Networks – can make a difference, as they possess the ability to learn from data and are thereby able to 'understand' complex processes better [52]. In the most recent papers concerning machine learning in the domain of ocean models, a trend could be observed of using neural networks in favor of more conventional ocean models. In this section we want to present some of the domains where neural networks are currently being researched to improve or assist existing practices.

Facts

- It is very timely expensive to locate drifting plastics or other floating waste in the vastness of the oceans. An autonomous way of detection would not only save resources but speed up the cleaning of the ocean [53].
- Ocean wave prediction can be helpful for energy production and planning of shipping routes [52].
- Sea level rises propose a threat to communities and coastal ecosystems [54].
- Ocean noise from human origin has impacted marine plants and the oceans soundscape [55].
- Through reflectance of the ocean, information about the health of the observed area can be derived.
- Ocean currents make the navigation of Unmanned Underwater Vehicles difficult [56].
- Through ocean subsurface temperature we can get information about the environmental conditions, the sea level as well as climate patterns [57].

Key Drivers

- Long Short Term Memory Neural Networks have shown great predictive performance in the applications of real time ocean current predictions, sea level predictions and also of wave heights. Outperforming conventional numerical models [58], established statistical models like ARIMA [59] or achieving very low reported errors (Root Mean Square Error in the range of 0.1 to 0.11) [56].
- A back-propagation neural network could be trained to obtain the subsurface ocean temperature from selecting optimal input data gained from satellite data, such that predictions could be improved [60].

Figure 2.4: These floating plastics could be detected by CNNs in the future such that it could be removed more efficiently by human helpers [62].

- Convolutional Neural Networks (CNN) could be trained which achieved very good results when it comes to classification application in the domain of ocean soundscape and floating plastic waste. With an accuracy of about 96% sound sources could be classified, which originated for example from sea life or humane origin [55]. In addition to that another CNN achieved an accuracy of 83% for plastic waste classification on satellite imagery [53].
- A so called "Neural Network Reflectance Prediction Model" was trained that models the ocean reflectance much more computational efficient than the simulations which are commonly used and also speed up algorithms which retrieve information about important properties like "colored dissolved organic matter (CDOM), sediments, phytoplankton, and pollutants in water" [61].

Challenges

- Collecting data in the deepest areas of the oceans could be challenging for humans and for robots. Human vehicles or robots that go deep below the ocean surface should be prepared to endure very high pressure, low temperatures and be able to communicate with the ship or station on the surface. While this is technologically possible, it remains financially difficult to achieve for most countries in the world.
- Due to the climate change, the environment could show much more fluctuating behavior which will make predictions even harder and will require

a constant data flow to be able to adapt or retrain models to possible changes.

- Machine Learning models like Neural Networks can consume large amounts of energy during training depending on the number of parameters and also require high quality hardware (i.e. GPUs) to be trained in reasonable time. This is not only expensive, but can also have effects on the environment due to the emissions from energy- and component-production.
- Most of the time, Neural Networks work like a black box of a model, where only input and output are really known. This is why underlying dynamics of the model will stay hidden and one would not gain more insight into the dynamics of the oceans when a new field would be directly explored with Neural Networks.

Impact

Producing accurate models of the oceans is very helpful for understanding how human actions affect our oceans when it comes to climate change and also for coping with the effects of climate change on the oceans, as for example rising sea levels. The rising incorporation of Neural Networks in oceanography has shown to outperform existing models which were already being used. Any kind of optimization in this domain will yield better predictions through which humans can adapt better to the ocean and also decrease errors which will have negative effects on the ocean. In addition to that, through the use of CNNs, completely new automated applications can be realised which help in cleaning up and conserving the ocean. Through the use of NNs, assessing ocean parameters, which yield information about the ocean's health, can be optimized and also become more precise. Overall, the versatility of Neural Networks can not only improve existing models but also create new fields of application when it comes to the ocean and therefore has the ability to be very impactful.

2.3 Conclusion

The sections above naturally do not constitute an exhaustive list of potential future trends in machine learning and data science connected to oceanography. Nevertheless, we believe that in the near foreseeable future, the applications for neural networks and other machine learning methods we mentioned will find much more widespread use within the marine science community compared to the current state-of-the-art.

Remote sensing technology will likely experience a widespread inclusion of specialised neural networks for detection- and classification-purposes, for instance when monitoring harmful algal blooms and oil spills, helping with climate change mitigation.

More autonomous seacrafts will likely be used, not only for further exploration of our oceans but also for facilitating the process of data acquisition and

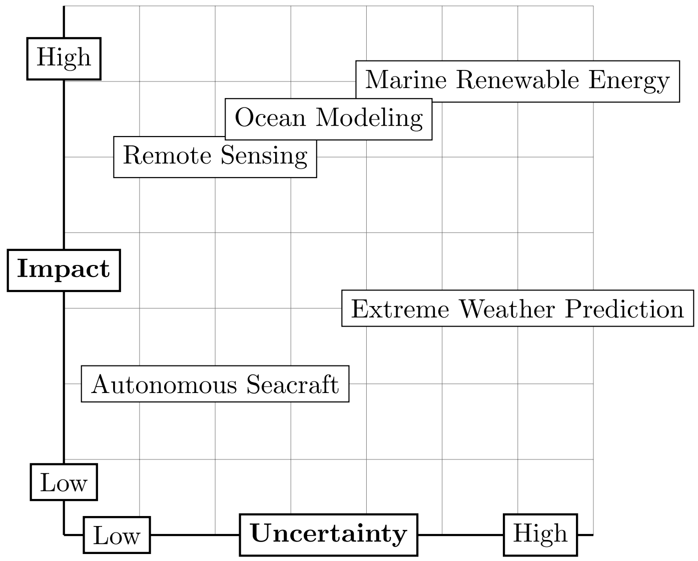

Figure 2.5: Driver matrix

freeing up labour force that could be used elsewhere in research.

The identification of marine renewable energy sources is another area of interest that we predict will see increased inclusion of machine learning, which will aid in solving some of the challenges scientists are currently facing, for instance in the development of offshore wind energy.

A drastically different field of study that we predict will see increased use of machine learning in the future is the forecasting of extreme weather conditions.

Finally, bringing all of the above together, more accurate models for ocean analysis in general could emerge as a result of improved neural networks.

References

[1] Camilo Mora et al. "How many species are there on Earth and in the ocean?" In: *PLoS biology* 9.8 (2011), e1001127.

[2] National Ocean Service. *How much of the ocean have we explored?* `https://oceanservice.noaa.gov/facts/exploration.html`. [Online; accessed: 29-August-2021].

[3] Sebastian Scher and Gabriele Messori. "How global warming changes the difficulty of synoptic weather forecasting". In: *Geophysical Research Letters* 46.5 (2019), pp. 2931–2939.

[4] Wikipedia. *Remote sensing — Wikipedia, The Free Encyclopedia*. `http://en.wikipedia.org/w/index.php?title=Remote\%20sensing&oldid=1033303173`. [Online; accessed 21-July-2021]. 2021.

[5] John Rogan and DongMei Chen. "Remote sensing technology for mapping and monitoring land-cover and land-use change". In: *Progress in planning* 61.4 (2004), pp. 301–325.

[6] Sander Veraverbeke et al. "Hyperspectral remote sensing of fire: State-of-the-art and future perspectives". In: *Remote Sensing of Environment* 216 (2018), pp. 105–121. ISSN: 0034-4257. DOI: `https://doi.org/10.1016/j.rse.2018.06.020`. URL: `https://www.sciencedirect.com/science/article/pii/S0034425718302980`.

[7] Wouter H. Maes and Kathy Steppe. "Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture". In: *Trends in Plant Science* 24.2 (2019), pp. 152–164. ISSN: 1360-1385. DOI: `https://doi.org/10.1016/j.tplants.2018.11.007`. URL: `https://www.sciencedirect.com/science/article/pii/S1360138518302693`.

[8] Terry P Hughes et al. "Global warming transforms coral reef assemblages". In: *Nature* 556.7702 (2018), pp. 492–496.

[9] Science On a Sphere. *Coral Reef Risk Outlook*. `https://sos.noaa.gov/catalog/datasets/coral-reef-risk-outlook/`. [Online; accessed 27-August-2021]. 2012.

[10] Ivor D Williams et al. "Leveraging automated image analysis tools to transform our capacity to assess status and trends of coral reefs". In: *Frontiers in Marine Science* 6 (2019), p. 222.

[11] National Oceanic and Atmospheric Administration. *What is a harmful algal bloom?* `https://www.noaa.gov/what-is-harmful-algal-bloom`. [Online; accessed 22-July-2021]. 2016.

[12] National Oceanic and Atmospheric Administration. *Why do harmful algal blooms occur?* `https://oceanservice.noaa.gov/facts/why_habs.html`. [Online; accessed 22-July-2021]. 2021.

[13] Water Australian Government: Department of Agriculture and the Environment. *Why we should recylce used motor oil*. `https://www.environment.gov.au/protection/publications/factsheet-why-we-should-recycle-used-motor-oil`. [Online; accessed 27-August-2021]. 2009.

[14] Isaac Segovia Ramırez et al. "Autonomous Underwater Vehicles and Field of View in Underwater Operations". In: *Journal of Marine Science and Engineering* 9.3 (2021), p. 277.

[15] Kalyan De et al. "Application of remotely sensed sea surface temperature for assessment of recurrent coral bleaching (2014–2019) impact on a marginal coral ecosystem". In: *Geocarto International* (2021), pp. 1–25.

[16] Mortimer Werther et al. "Meta-classification of remote sensing reflectance to estimate trophic status of inland and nearshore waters". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 176 (2021), pp. 109–126.

[17] Mitchell B. Lyons et al. "Mapping the world's coral reefs using a global multiscale earth observation framework". In: *Remote Sensing in Ecology and Conservation* 6.4 (2020), pp. 557–568.

[18] Catherine Jordan et al. "Using the Red Band Difference Algorithm to Detect and Monitor a Karenia spp. Bloom Off the South Coast of Ireland, June 2019". In: *Frontiers in Marine Science* 8 (2021), p. 638889.

[19] Ziyi Suo et al. "Ultraviolet remote sensing of marine oil spills: a new approach of Haiyang-1C satellite". In: *Optics Express* 29.9 (2021), pp. 13486–13495.

[20] Stefania Magrı et al. "Oil Spill Identification and Monitoring from Sentinel-1 Sar Satellite Earth Observations: a Machine Learning Approach". In: *Chemical Engineering Transactions* 86 (2021), pp. 379–384.

[21] Marcos RA Conceição et al. "SAR Oil Spill Detection System through Random Forest Classifiers". In: *Remote Sensing* 13.11 (2021), p. 2044.

[22] Agathe Puissant et al. "Inversion of Phytoplankton Pigment Vertical Profiles from Satellite Data Using Machine Learning". In: *Remote Sensing* 13.8 (2021), p. 1445.

[23] Claudia Glover. *Intelligent machines are fighting plastic pollution in the world's oceans.* https://techmonitor.ai/leadership/innovation/ai-digital-innovations-marine-plastic. [Online; accessed: 30-August-2021]. 2020.

[24] Karolina Zwolak et al. "The autonomous underwater vehicle integrated with the unmanned surface vessel mapping the Southern Ionian Sea. The winning technology solution of the Shell Ocean Discovery XPRIZE". In: *Remote Sensing* 12.8 (2020), p. 1344.

[25] Jing Yan et al. "AUV-Aided Localization for Internet of Underwater Things: A Reinforcement-Learning-Based Method". In: *IEEE Internet of Things Journal* 7.10 (2020), pp. 9728–9746. DOI: 10.1109/JIOT.2020.2993012.

[26] Philip Solaris. *The Green Desk: Using Autonomous Seacraft to Tackle Overfishing with Philip Solaris.* https://95bfm.com/bcast/the-green-desk-using-autonomous-seacraft-to-tackle-overfishing-with-philip-solaris-april-28-2020. [Online; accessed: 30-August-2021]. 2020.

[27] Pacific Waves. *Autonomous seacraft to police Pacific fisheries.* https://www.rnz.co.nz/international/programmes/datelinepacific/audio/2018753768/autonomous-seacraft-to-police-pacific-fisheries. [Online; accessed: 30-August-2021]. 2020.

[28] Melanie Gömmel. *#StopPlasticPollution: Warum fischen wir das Plastik nicht aus dem Meer?* https://blog.wwf.de/stopplasticpollution-warum-fischen-wir-das-plastik-nicht-aus-dem-meer/. [Online; accessed: 30-August-2021]. 2019.

[29] Enrica Zereik et al. "Challenges and future trends in marine robotics". In: *Annual Reviews in Control* 46 (2018), pp. 350–368.

[30] Andrzej Felski and Karolina Zwolak. "The Ocean-Going Autonomous Ship—Challenges and Threats". In: *Journal of Marine Science and Engineering* 8.1 (2020). ISSN: 2077-1312. DOI: 10.3390/jmse8010041. URL: https://www.mdpi.com/2077-1312/8/1/41.

[31] Samantha Jordan. "Captain, My Captain: A Look at Autonomous Ships and How They Should Operate Under Admiralty Law". In: *Indiana international & comparative law review* 30(2) (2021), pp. 283–317. ISSN: 2169-3226. DOI: 10.3390/jmse8010041.

[32] *Marine Renewable Energy*. https://www.mdpi.com/topics/marine. [Online; accessed 22-July-2021].

[33] David Rolnick et al. *Tackling Climate Change with Machine Learning*. https://arxiv.org/pdf/1906.05433.pdf. [Online; accessed: 27-August-2021]. 2019. arXiv: 1906.05433 [cs.CY].

[34] Deutsche WindGuard GmbH. *Status of Onshore Wind Energy Development in Germany Year 2020*. https://www.wind-energie.de/fileadmin/redaktion/dokumente/dokumente-englisch/statistics/Status_of_Onshore_Wind_Energy_Development_in_Germany_-_Year_2020.pdf. [Online; accessed 22-July-2021].

[35] Walter Musial, Sandy Butterfield, and Bonnie Ram. "Energy from offshore wind". In: *Offshore Technology Conference*. OnePetro. 2006.

[36] Alistair GL Borthwick. "Marine renewable energy seascape". In: *Engineering* 2.1 (2016), pp. 69–78.

[37] Markus Schmitt. *Machine Learning for Energy Generation*. https://towardsdatascience.com/machine-learning-for-energy-generation-302069a942f. [Online; accessed 22-July-2021]. 2020.

[38] David Pugh and Philip Woodworth. *Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes*. Cambridge University Press, 2014.

[39] Jens Peter Kofoed et al. "Prototype testing of the wave energy converter wave dragon". In: *Renewable energy* 31.2 (2006), pp. 181–189.

[40] Adrian Stetco et al. "Machine learning methods for wind turbine condition monitoring: A review". In: *Renewable energy* 133 (2019), pp. 620–635.

[41] Xiuxing Yin and Xiaowei Zhao. "Big data driven multi-objective predictions for offshore wind farm based on machine learning algorithms". In: *Energy* 186 (2019), p. 115704.

[42] Dripta Sarkar, Michael Osborne, and Thomas Adcock. "A Machine Learning Approach to the Prediction of Tidal Currents." In: *The 26th International Ocean and Polar Engineering Conference*. OnePetro. 2016.

[43] Seyed Milad Mousavi et al. "Deep Learning for Wave Energy Converter Modeling Using Long Short-Term Memory". In: *Mathematics* 9.8 (2021), p. 871.

[44] Erich Westendarp. *Windrad Offshore Wattenmeer.* https://pixabay.com/photos/pinwheel-offshore-wadden-sea-watts-1317817/. [Online; accessed 27-August-2021]. 2013.

[45] Marius Årthun et al. "Skillful prediction of northern climate provided by the ocean". In: *Nature communications* 8.1 (2017), pp. 1–11.

[46] Clifford Broni-Bedaiko et al. "El Niño-Southern Oscillation forecasting using complex networks analysis of LSTM neural networks". In: *Artificial Life and Robotics* 24.4 (2019), pp. 445–451.

[47] Jennifer A Hanafin and Peter J Minnett. "Measurements of the infrared emissivity of a wind-roughened sea surface". In: *Applied optics* 44.3 (2005), pp. 398–411.

[48] Li Wei, Lei Guan, and Liqin Qu. "Prediction of Sea Surface Temperature in the South China Sea by Artificial Neural Networks". In: *IEEE Geoscience and Remote Sensing Letters* 17.4 (2019), pp. 558–562.

[49] Changjiang Xiao et al. "A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data". In: *Environmental Modelling & Software* 120 (2019), p. 104502.

[50] Lihao Wang, Yu Jiang, and Hong Qi. "Marine Dissolved Oxygen Prediction With Tree Tuned Deep Neural Network". In: *IEEE Access* 8 (2020), pp. 182431–182440.

[51] Martin Mäll et al. "Modeling Parameters and Impacts of Future Cyclones: South-East Asian and Northern European Case Studies". In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2019, pp. 9346–9349.

[52] Ranran Lou et al. "Application of machine learning in ocean data". In: *Multimedia Systems* (2021), pp. 1–10. ISSN: 1432-1882. DOI: 10.1007/s00530-020-00733-x. URL: https://link-springer-com.eaccess.ub.tum.de/article/10.1007/s00530-020-00733-x#ref-CR31.

[53] Mattis Wolf et al. "Machine learning for aquatic plastic litter detection, classification and quantification (APLASTIC-Q)". In: *Environmental Research Letters* 15.11 (2020), p. 114042. ISSN: 1748-9326. DOI: 10.1088/1748-9326/abbd01. URL: https://iopscience.iop.org/article/10.1088/1748-9326/abbd01.

[54] Michael Oppenheimer et al. "Sea Level Rise and Implications for Low Lying Islands, Coasts and Communities". In: *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. 2019. URL: https://www.researchgate.net/publication/336134630_Sea_Level_Rise_and_Implications_for_Low_Lying_Islands_Coasts_and_Communities.

[55] B. Mishachandar and S. Vairamuthu. "Diverse ocean noise classification using deep learning". In: *Applied Acoustics* 181 (2021), p. 108141. ISSN: 0003-682X. DOI: 10.1016/j.apacoust.2021.108141. URL: https://www.sciencedirect.com/science/article/pii/S0003682X21002358.

[56] Alexandre Immas, Ninh Do, and Mohammad-Reza Alam. "Real-time in situ prediction of ocean currents". In: *Ocean Engineering* 228 (2021), p. 108922. ISSN: 0029-8018. DOI: 10.1016/j.oceaneng.2021.108922. URL: https://www.sciencedirect.com/science/article/pii/S0029801821003577.

[57] National Centers for Environmental Information. *GCOS Ocean Physics ECV - Subsurface Temperature*. https://www.ncdc.noaa.gov/gosic/gcos-essential-climate-variable-ecv-data-access-matrix/gcos-ocean-physics-ecv-subsurface-temperature. [Online; accessed: 27-August-2021].

[58] Song Gao et al. "A forecasting model for wave heights based on a long short-term memory neural network". In: *Acta Oceanologica Sinica* 40.1 (2021), pp. 62–69. ISSN: 1869-1099. DOI: 10.1007/s13131-020-1680-3. URL: https://link.springer.com/article/10.1007/s13131-020-1680-3.

[59] Abdul-Lateef Balogun and Naheem Adebisi. "Sea level prediction using ARIMA, SVR and LSTM neural network: assessing the impact of ensemble Ocean-Atmospheric processes on models' accuracy". In: *Geomatics, Natural Hazards and Risk* 12.1 (2021), pp. 653–674. ISSN: 1947-5705. DOI: 10.1080/19475705.2021.1887372.

[60] Hao Cheng, Liang Sun, and Jiagen Li. "Neural Network Approach to Retrieving Ocean Subsurface Temperatures from Surface Parameters Observed by Satellites". In: *Water* 13.3 (2021), p. 388. DOI: 10.3390/w13030388. URL: https://www.mdpi.com/2073-4441/13/3/388.

[61] Lipi Mukherjee et al. "Neural Network Reflectance Prediction Model for Both Open Ocean and Coastal Waters". In: *Remote Sensing* 12.9 (2020), p. 1421. DOI: 10.3390/rs12091421. URL: https://www.mdpi.com/2072-4292/12/9/1421.

[62] Naja Bertolt Jensen. *school of fish in water*. https://unsplash.com/photos/BJUoZu0mpt0. [Online; accessed 02-September-2021]. 2021.

Chapter 3

Artificial Intelligence for Solar Energy

DAI, BAIQI
DEMIRDÖGEN, CAN
FU, YI;
MEI, GAO
PRODAN, MARIA-EFIA
SELMANI, EROLL;
UTKU, AYVAZ
SHAH, MENAYL
WANG, HUIYU

Abstract

With the increase of global warming, efforts are underway to control this issue and reduce the carbon footprint as much as possible. One of the many of such efforts is to shift from fossil fuel-based energy to clean renewable energy resources such as wind, solar, etc. and integration of Artificial Intelligence (AI) with renewable energy sources to make the deployment of these technologies more efficient. This paper focuses on photovoltaic-based (PV) solar energy technology and discusses various trends for the application of AI in solar energy technology. The general background, facts, challenges and impact of different trends such as automatic defect and fault detection, energy sharing, organic solar cells research, and estimating solar energy potential on house roofs virtually are discussed. Following these trends, a smarter, efficient and clean deployment of PV-based technology is ensured with a medium impact on the global carbon footprint with the potential to grow even bigger.

3.1 Introduction

Rising issues about climate change, the health effects of air pollution, energy security and access, as well as volatile oil prices in recent decades, have necessitated the development and usage of low-carbon technology alternatives such as renewable energy. Over the years, solar PV has been one of the most innovative renewable energy technologies. By the end of 2018, the total installed capacity of solar PV has surpassed 480 GW globally, making it the second-largest renewable energy source behind wind [1].

Renewable energy deployment has increased at a rapid speed recently, breaking records and overtaking yearly conventional power capacity expansions in several locations. Solar PV power installations have dominated the renewable business for many years, outperforming all other renewable technologies. According to a new study, solar PV power installations might nearly double in size over the next ten years, reaching 2 840 GW by 2030 and 8 519 GW by 2050. This means that the total installed capacity in 2050 will be about eighteen times more than it was in 2018 [1] (see Figure 3.1).

However, this very promising development also causes multiple challenges that need to be mastered in the upcoming years. The growing number of PV modules makes the monitoring, maintenance and repair increasingly complex. Significant cost reductions are driven by government actions, such as deployment policies, finance, and other policies that have encouraged the PV installations on private rooftops, which have led to a breakthrough in renewable capacity increases in recent years. The calculation of the solar potential of a single rooftop, however, takes between one hour to two full days which is extremely unfavorable in view of the cost. The development in this area is thus slowed down consequently automation is desperately needed [1].

In order to cope with the increasing PV demand, millions of new solar cells are going to be produced in factories that are predominantly located in southeast Asia. It is intuitive that resources must be spent on research and development of more efficient and organic Solar cells in order to reduce the environmental pollution that occurs during production [1].

There are a lot of different approaches in order to overcome the problems mentioned above. One of these approaches, which is the most promising, is the usage of machine intelligence. Various machine-learning algorithms are already in use in renewable-energy projections, according to numerous research. MI enables high precision forecasting and improves the efficiency of the energy system. Due to its high versatility, it is a suitable tool to tackle the mentioned challenges.

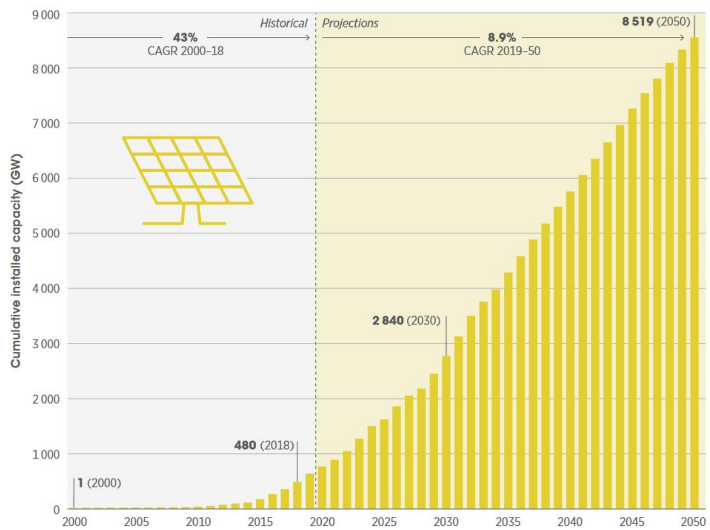

Figure 3.1: Cumulative solar PV capacity growth[1]

3.2 Trends

The following section describes various trends in the photovoltaic sector. Firstly, automatic fault detection of solar power plants is presented. This should contribute to the effective and cost-efficient mitigation of solar panel failures. Next, the area of Energy Sharing is considered in more detail. The third trend deals with research in the field of organic solar cells. Developments in this area can contribute to a significant reduction in the price of solar cells. In the end, the solar energy potential of housing roofs is described.

3.2.1 Automatic defect and fault detection

The solar power plant surfaces will increase significantly in the next decade. This is implying many efforts in the maintenance of the solar panels. The largest solar power plants in the world are able to produce hundreds and thousands of millions of watts. It is very clear that manual inspection of solar panels for surface defects is not efficient. Furthermore, some of the defects are even harder to be detected by the human eye. Also, the best experts in the field fail sometimes to detect in time the irregularities that can exist on the surfaces of the photovoltaic modules. This is why computer vision and learning techniques are needed: to improve the production quality and the efficiency of solar cells.

General background

Photovoltaic modules can experience thermo-mechanical stress during production and usage. This stress can induce defects, for example, cracks, chromatic

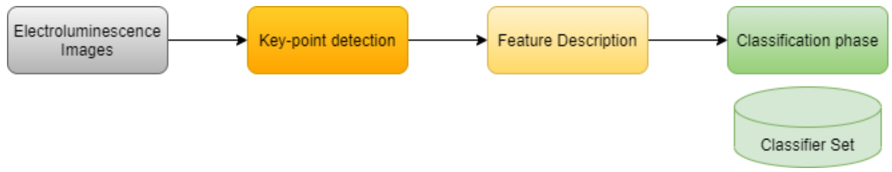

Figure 3.2: The process of detecting key-points in an image and classify the defects with machine learning methods

aberration and yellowing, which can affect the power output [**fatiguedegradation**]. These defects are highly probable to cause faulty conditions in the modules: mismatch fault, open circuit fault, bridging fault, earth fault, etc. One example of a failure in the modules is the defective cells, which might not have a uniform distributed current density due to the presence of shunts that may be created by material defects. This will result in a lower output of a PV module. One of the most common reasons for the loss in the PV modules output is generated by cracks.[**fatiguedegradation**]

Facts

- Drones inspection and data collection

There are various techniques to inspect PV modules. Each technique has several advantages and limitations and appears to be tailored to particular defects. Drones are used for inspecting PV systems. They are often equipped with RGB optical and thermal camera systems for inspections. Detectors commonly utilized in PV infrared inspection are micrometer detectors. They fall into the thermal category where the detector operates on the basis of heating by incident infrared radiation [2]. Measurements may be accomplished in two ways. The first is called quantitative, in which, genuine temperature values of items are obtained. The second is qualitative. It gives the relative temperature compared to different components with inside the inspected object. As mentioned the drones are equipped with two camera systems. Drones use the pictures of the thermal camera system for scanning the area for autonomous flight. When heat radiation from the surface is detected, the drone knows it is a solar module. In this case, the flight area can be defined. The next step is to analyze the scanned images. Using image processing the several defects of PV modules can be detected.

- Detection of defects

The inspection of PV modules is firstly done by the acquisition of images. A very efficient solution is represented by autonomous drones. The acquired data set is latest streamed into the pre-processing step, done by computer vision techniques, and into the classification step, in which different classification Machine Learning Algorithms are used for surface defect detection of the modules. The choice of the algorithm is depending on the type of defect that it is targeted to be detected.

- Fault detection

Another method to detect a potential fault in the system is using supervised learning (Forest Regression Algorithm) to predict the output of a photovoltaic module, where the input is represented by environmental data, such as air temperature, air humidity, air pressure and date. A relative fault is found by comparing the real output from PV system with the expected output. A time horizon has to be defined e.g. one week, one month. The error is calculated using:

$$error = \frac{avg(out_{expected} - out_{measured})}{out_{measured}}$$

Then the error is compared with a threshold e.g(50%). If the error exceeds the threshold, then it indicates a fault.[3]

Key drivers

- Autonomous drones and camera systems: intelligent image acquisition of solar panels for further processing.
- Computer vision and machine learning methods [4]: accelerates the efficiency and accuracy of defects detection and increases the quality and productivity of the solar panels.
- Fault detection: prediction of a fault caused by a defect using environmental data.

Challenges

- Environmental Conditions for drone flight. Before beginning any drone inspection, the onsite environmental condition needs to be checked. Accurate, high-quality thermal data requires specific weather conditions, including clear skies and clear weather or in the worst case clear cloudy skies. For every drone, solar farm inspection sufficient irradiance needs to be confirmed. Also, humidity should be less than 60%, and the wind speed below 15 MPH [5].
- Improvement of time horizon in the fault detection. When the threshold is 50%, it's only possible to detect the fault correctly when the time horizon is bigger than two weeks. A smaller time horizon leads to the problem that non-existing faults are increasingly reported. This problem can be compensated by increasing the threshold. However, after that more actual faults are not reported. On the other hand, increasing the time horizon also leads that our fault could not be found in a short time. Since it's important to find the appropriate value for the time horizon and the threshold, the mentioned method to detect faults has to be improved, in order to find faults in a shorter time horizon.

- Imbalanced data set[4] . If the accuracy is used as a performance measure, then this may result in a false sense of high performance. The common performance metrics in the research community are precision, recall, and F-score in this case.

Impact

The maintenance of the PV system is an important task to improve the output and the efficiency of the system. Computer vision, machine learning and autonomous drones can bring efficiency in detecting different causes that could decrease the output of a solar panel. Moreover, these high computational methods can have a huge impact in reducing manual work, human error and decrease the human resource cost.

3.2.2 Energy Sharing

Cars, smart phones, holiday homes - in the "sharing economy" many things are no longer bought, but shared. In the future, this should also apply to electricity from renewable energies.

General background

A power-sharing, which enables a common, decentralized consumption of locally produced renewable electricity, could allow an economical operation of all these PV Roof systems.

- Energy community model

  The energy surplus, energy deficit for each user can be evaluated, also the energy produced within the energy community[6]. If there is not any energy storage system, the amount of energy that can be managed in the community is equal to the energy that is shared in real-time. If a centralized energy storage is considered, the community can manage more energy.

- Collective investment evaluation

  In order to evaluate the feasibility of a specific investment for the energy community, a large number of variables must be considered, such as the initial investment, the operational costs, the savings obtained self-consuming/sharing energy and the administrative cost, etc.

- Machine learning approach

  Based on machine learning methods, it is possible to better construct and optimize solar power generation forecasting models and assess environmental impacts.

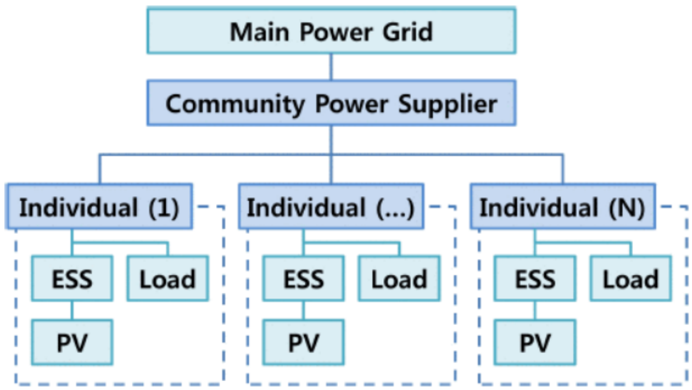

Figure 3.3: Block diagram of the overall energy sharing smart community

Some papers propose the method for a smart Residential Energy Management System (REMS) using Machine Learning [7]. Specifically, the proposed system (REMS) effectively switches pre-prioritized possible loads without limiting consumption, between the grid and renewable energized local storage with rooftop solar at the residential premises, using Machine Learning algorithms.

Another supervised machine learning approach is to predict and schedule the real-time operation mode of the next operation interval for residential PV/Battery systems controlled by mode-based controllers [8]. Which may significantly reduce the cost of implementing such systems.

In addition, machine learning algorithms also could model the entire energy community as a whole, rather than each presume separately, with the goal of optimizing the energy sharing and balance at the community level.

Facts

- The European Commission took interest in Energy Communities (EC) as a tool that enables European citizens to take part in the so-called clean energy transition [6].

- The energy sharing framework assumes the existence of individuals that own their renewable power resources within the smart community. A photovoltaic power generation system (PV) with an energy storage system (ESS) is the most common renewable resource that individuals could install in their territories [9].

- Administrative regulations, technical regulations and electricity laws in these buildings are a major obstacle to the widespread adoption of on-site PV power, as they make implementation more difficult and significantly reduce the financial attractiveness for owners. In particular, the millions of fully functional smaller PV systems on private houses could hardly be operated economically and face an uncertain future with the risk of being switched off.

Key Drivers

- Further development of AI technology

  We need higher-precision AI forecasting models to meet people's demand for solar power forecasting in the future. When AI technology is further developed, the use of AI can make solar forecasts more accurate and valuable. It helps to improve the data rigor of grid operation. Because AI can help operators make smarter, faster, and more data-driven assessments of how their power generation meets power demand.

  In addition, AI can also balance power supply and demand, whether it is power surplus or vacancy, AI can prevent this from happening.

- The development of photovoltaic modules

  As we all know, photovoltaic modules are a very important link in the shared solar economy. The development of photovoltaic modules also determines the upper limit of the photovoltaic industry. Conversion efficiency is an indispensable consideration in the development of photovoltaic modules in recent years.

  The development of solar cells is also a top priority. At present, the conversion efficiency of mainstream monocrystalline silicon solar cells exceeds 22%. If high-efficiency and low-cost solar cells can be further developed in the future, the conversion efficiency of the entire photovoltaic module will be further improved.

- The formation of a unified platform

  The formation of a unified platform [10] marks the formation and improvement of the solar energy sharing economy. It will solve the problem of inconsistent pricing and management of many private solar companies. And it will become a strong guarantee for the installation, cleaning, maintenance, and dismantling of the solar energy industry. The unified platform that will emerge in the future will provide investors from all over the world with investment channels, allowing people to share the dividends of solar energy even without a roof [11]. The safe and stable large-scale platform can attract more investors, and further accelerate the installation plan of rooftop solar panels. The larger the industry, the lower the cost and the higher the profit. In the near future, the solar energy sharing economy will go to the world.

Challenges

- The promotion of rooftop installation of solar panels requires people's conceptual acceptance. At present, roof installations must consider roof load issues and slow cost recovery issues. The cost recovery is about 50% in five years.

- For the platform, large-scale capital investment must be required in the early stage. In the face of today's investment field, sustainable energy investment is still a minority. If the capital needs of the platform can be passed, the follow-up gains are only a matter of time.

- As everyone knows, technology is updated very quickly. It is impossible to predict the form of solar energy utilization in the future. If the new technology can abandon solar panels, the recycling and disposal of old panels will be a huge problem. How to manage decommissioned photovoltaic panels will be an issue that people need to consider at all times.

Impact

It is foreseeable in the future that, driven by global investment, solar panels installed on roofs will be popularized in a large area, and the distributed power generation system will be further improved. People can more easily obtain clean and environmentally friendly electricity, and the problem of global warming will also be alleviated.

3.2.3 Organic Solar Cell Research

Over the years the interest in solar panels has been increasing and the price of solar energy has been decreasing [12]. However, solar panel equipment and installation have relatively high prices, which makes solar energy inconvenient in comparison to other less environmentally friendly energy resources.

Solar cells are the smallest components of a PV system, which converts sunlight into electricity. As such the weight, efficiency, and flexibility of the solar cells inside the PV system affect the quality and the price of the PV system. Currently, the most efficient solar cells in the market are the silicon solar cells. However, silicon solar cells lack flexibility and are relatively expensive. A solution to this problem can be organic solar cells, which are flexible, light, and inexpensive to produce. However, a large disadvantage of organic solar cells is their efficiency, which is lower than the state-of-the-art silicon solar cells.

General background

- General research direction

The efficiency of organic solar cells is dependent on the choice of electron donors and electron acceptors [13]. Although, many donors and acceptors were tested, due to the large degrees of freedom in organic molecules, there is still an extremely large number of possibilities left. Therefore, the research in the search of efficient electron donor-acceptor pairs is being done by trial-and-error but this is time-consuming. Moreover, due to the chemical structure of organic molecules, which are extremely complex, it is hard to forecast if an electron donor-acceptor pair will have high efficiency before synthesizing and testing the said pair. With this method, the goal is to find patterns in acceptor and donor pairs that result in high efficiency.

- Properties of donor and acceptors

In order to evaluate the feasibility of a specific investment for the energy community, a large number of variables must be considered, such as the initial investment, the operational costs, the savings obtained self-consuming/sharing energy and the administrative cost, etc.

- Machine learning approach

Instead of doing trial and error, an idea is to use Machine learning to forecast the power conversion efficiency (PCE) of the donors and acceptors. Firstly, we can estimate the PCE without doing time-consuming synthesizing and testing. Secondly, if the machine learning model is accurate with correctly distributed data machine learning can help us find patterns of electron donors and acceptors. Machine learning models use different types of properties. Many papers use different properties of donors and acceptors and different machine learning models [13]. One of these machine learning models is CasSVM[14]. This model uses molecular descriptors of donor and acceptor to estimate the Short-circuit current density, Open-circuit voltage, Fill factor and PCE. The Short-circuit current density, Open-circuit voltage and Fill factor are estimated by three different Support Vector Machines (SVM), which takes molecular descriptors as input. Likewise, PCE is estimated by another SVM, which takes Short-circuit current density, Open-circuit voltage and Fill Factor as input.

Facts

Some promising models were developed in the pursuit of finding patterns, that result in high PCE. Here we share a single model per input type, which is used to find patterns and their results:

- Molecular Descriptors (CasSVM)
  CasSVM resulted in a 0.96 Coefficient of determination between its inputs and the PCE. However, a pessimistic side of this model is the fact that it was trained with a relatively small data set, which has 161 data points.[14]

- Molecular Fingerprints (KRR)
  KRR model resulted in a 0.78 Correlation coefficient between its inputs and the PCE. Like CasSVM this model also has a small data set consisting of 320 data points.[15]

- Microscopic Properties (GBRT)
  GBRT resulted in a 0.80 Correlation coefficient between its inputs and the PCE. Like the previous entries, this model also has a small data set consisting of 300 data points.[16]

- Energy Levels (RF)
  RF resulted in a 0.80 Coefficient of determination between its inputs and the PCE. This model had a data set of 135 data points.[17]

Key Drivers

- ML analysis steps

Sample collection: Collect data from experiments or calculations and process the data, such as data conversion or data segmentation, and divide the data into training sets and test sets. The performance of the model will be affected by data segmentation. A reasonable ML model should have more data, and the quantity and quality of data will also greatly affect the performance of the model. Data preparation and processing: data processing can effectively improve the quality of modeling. PCA, LDA and ICA are commonly used analysis methods. Modeling: Due to the complex relationship between parameters and performance, modeling will facilitate data analysis and improve the accuracy of the model. Model evaluation: Statistical analysis is usually used to test the performance of the model, using root mean square error (RMSE) and determination coefficient (R2).

Challenges

- Descriptor selection

Descriptor type determines the ML model prediction ability, while the choice of descriptor based on the properties of the target, before the ML model experiments first should choose the most appropriate descriptors, the selected descriptors should be easy to calculate, in preparation for the screening of appropriate equipment, common descriptors include molecular descriptors, molecular fingerprints. Molecular descriptors have different dimensions, from one-dimensional to three-dimensional. Descriptors of different dimensions have different meanings. One-dimensional descriptors describe the types of chemical fragments, two-dimensional descriptors describe topological chemical molecular features, and three-dimensional descriptors describe the geometric data of molecules. Molecular descriptors are easy to calculate quickly.

- Data infrastructure

Data can be used for material screening of ML models, but the structure of data is often complex, and ML model prediction depends on the difference of data structure. However, building a complete ML model requires a lot of data.

There are hundreds of data points in a solar organic cell, and the accuracy of the model increases as the data increases. Large computations often require data simplification, which requires a balance between the available data and the predictive power of the model.

Choosing the right organic semiconductor and finding efficient materials for solar cells is difficult because it is difficult to judge a material from its chemical structure alone, and organic solar cell materials remain challenging to achieve.

Impact

More flexible and inexpensive solar cells can result in larger use of PV systems worldwide. Consequently, the solar energy would be used more often by public, which helps the fight against the climate change by reducing the greenhouse gas emission around the world. However, as mentioned before there are too many degrees of freedom in properties of electron donors and acceptors, which makes it time-consuming to find the patterns. Therefore, although the foundation of organic solar cells, which are more efficient in comparison to silicon solar cells, can have a large impact but the it is uncertain.

3.2.4 Estimating Solar Energy Potential On House Roofs

The investment of big companies such as Total and Google to develop AI-based solutions for the solar industry shows the future potential of AI integration in this domain. A tool called Solar Mapper, which aims to accelerate the deployment of solar panels for individuals by providing an accurate and rapid estimate of the solar energy potential of their homes is created. 'Solar Mapper is contributing to the Group's ambition to become a world leader in the production of renewable energies, toward getting to net-zero emissions by 2050 together with society' according to Total and Google. Solar Mapper makes use of AI that provides better results than current tools.

General background

The dream of an electrically self-sufficient home has become achievable since the usage of solar panels in the private sector was made possible. Big metropolises in Europe and the US began to realize the opportunities this concept can create and started to equip rooftops of buildings with PV panels.

The figure shows that especially the largest cities in Germany, such as Munich, Frankfurt and Berlin, only exploit the PV potential to a very small extent. However, the report also mentions that the overall development particularly in the area of new buildings, moves upwards to higher utilization, although very slowly.

The main issue that is currently holding back the development of this trend is that identifying and estimating the PV potential of an individual roof takes between 1 hour to 2 full days. According to industry calculations, this time-consuming process takes up to 30-40% of total project costs making it extremely financially unprofitable [18].

Facts

Currently there are various techniques to detect roofs and identify the rooftop area suitable for PV usage [19]:

- Constant-value method

This method just assumes a certain percentage of rooftop area as available for PV usage. Gagnon et al.[19] derive from the literature that approximately 22-27% of residential roofs and 60-65% of commercial roofs are suitable in this scenario. These percentages are then simply multiplied with the total building stock. The result of this calculation is imprecise as several influencing factors are not taken into account, nevertheless, it allows a very quick rough estimate for researchers.

- Manual selection

This methods evaluate a building individually in terms of capable PV area. A wide range of tools (aerial photography, Google earth etc.) is used for this purpose. Manual selection is the most precise estimation method but also the most time-consuming while not being scalable as well.

- Geographic information system (GIS)

GIS-based methods are the most reliable for a large-scale estimate of appropriate rooftop space. These approaches are more exact than constant-value methods, and they can manage far bigger data sets compared to manual selection.

In the following, the individual steps of the roof recognition are now briefly described. Initially, high-resolution satellite images are taken of the desired region. These images are then divided into equally sized squares which are preprocessed by creating thousand of smaller tiles. A team of researchers then annotates thousands of rooftops in hundreds of tiles. This way perfectly labeled data sets for Supervised Learning algorithms are generated.

The primary objective is to use Computer Vision models to detect rooftops in a given image. In addition to this, it is also important to determine the structure of a roof (Flat-, Hip-, Shed-roof), this creates a segmentation problem. In order to calculate the rooftop's effective area, obstacles like pipes, chimneys and vents must be subtracted from the whole. Thus obstacle identification is necessary. In [18], [20] the lack of labeled data for this specific case is explained. As a consequence, the researchers used an edge detection unsupervised approach and contours. Obstacles were recognized from the plain area to a large extent by placing a threshold on contour colors. Within the pictures the obstacles are then outlined in red and the roof area in green.

Key drivers

- Governmental funding

Multiple administrations around the globe begin to realize the potential of this trend and start to promote the expansion with several resources. Turkey for example introduced a law in 2019 allowing real or legal persons to produce electricity without the need of a license and sell their excess production to the grid. This policy is likely to result in an increase in distributed energy generation. As a result, raising awareness about Turkey's rooftop PV potential has become a priority [21].

Germany, the country that already ranks first in per capita photovoltaic installations in the world and fourth in cumulative capacity, is taking it a step further. In the early months of 2021, the state of Baden-Württemberg published a new draft law that requires the roofs of all new buildings to be equipped with PV systems. The legislator hopes that this will lead to a significant increase in the utilization of building roof areas [22][23].

- Increasing use of machine learning in the field of renewable energies

Solar light, solar heat, and wind are all incredibly varied renewable energy sources, and the subsequent variations in generation capacity can cause power system instability. These challenges motivated the usage of machine learning algorithms to improve the handling of energy consumption and generation. The use of machine intelligence has become indispensable in this field. Since machine intelligence is still far from its technological peak, the possibilities that will arise in the future in interaction with renewables are promising [24]. Large companies such as Google and projects like Solar Ai and Project Sunroof, are already investing a high amount of resources in the development of PV tools.

Challenges

- Necessity of high resolution satellite imaginary

The quality and zoom capability of satellite images of housing roofs is vastly dependent on the geographical location. In this context, geographic location does not necessarily mean the influence of mountains or cloud cover for example, but instead simply which continent the location is. In [25] Kumat et al. describe the difficulties of obstacle recognition when using satellite imagery in Indian regions. Compared to satellite images from the US, the zoom levels are lower and the 3D mapping capability is also missing. Without the necessary sharpness and clearness, it is quite difficult to mark out the optimal area for PV panels even for the human eye. Kumar et al.[25] explain that as a consequence the Adaptive Canny Edge Detection Algorithm, which was used by the group of researchers, provided many false-positive results which made the satellite images almost useless. The availability of high-resolution images is therefore essential. Considering this fact, the quality of satellite images must improve in order for this trend to be implemented in other countries besides the US as well.

Impact

With the help of detecting tools the process of suitable area for PV estimation is made significantly cheaper and less time consuming for households. This removes the last major barriers that hinder the full exploitation of the roof's potential. The availability of freely usable software from big giants like Google brings the whole subject even further into the light of publicity and fuels the trend.

There is also an interesting side effect of solar panels on rooftops, particularly in dense metropolitan areas. In [26] a group of Indian researchers analyzed the impact on the building cooling load of such setups. The researchers concluded that shadows created by the solar panels can reduce the cooling load of buildings by up to 90% and thus also significantly reduce the use of air conditioning in hot geographical areas like India.

3.3 Conclusion

Firstly, these trends have a great impact on PV system, such as more accurate prediction, human cost reduction. However, they are also facing a lot of challenges, such as unconvincing prediction, high cost, long time consuming.

Secondly, the uncertainty of these trends depends on for example new laws, the improvement of critical technology. With the support of law, the spread of trend would be accelerated. If the critical technology can be improved, the trend would face less barrier.

A trend matrix is shown in Fig.7.3 to give a whole aspect of these four mentioned trends.

Trend 1: Automatic defect and fault detection in PV system.
Trend 2: Energy Sharing.
Trend 3: Organic Solar Cell Research.
Trend 4: Estimating Solar Energy Potential On House Roofs Virtually.

The first trend about automatic fault and defect detection for PV system has a low uncertainty since it can be already practically applied, and drones are commonly used . The impact is medium. Although automatic maintenance can save human cost, the environmental condition restriction is still a difficulty. Also the mentioned fault detection method for the maintenance costs too long time, which needs to be improved.

The second trend about energy sharing has the lowest impact, because of the lack of new research and the slow development of customers. The price of electricity would drop. However, this trend spreads with a limited tempo due to high investment, slow cost recovery and the problem with conceptional acceptance. It has the lowest uncertainty since it can already be implemented in daily life.

The third trend about organic solar cells has not only the greatest impact,

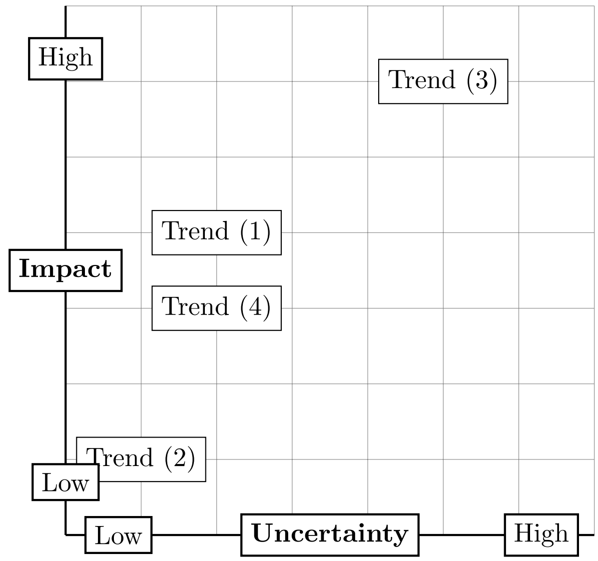

Figure 3.4: Driver matrix

but also the highest uncertainty. Although a big amount of research shows great potential in lowering the price of solar panels by finding the necessary patterns to increase the efficiency of organic solar cells, showing a promise for the future. However, there are still challenges such as high dimensionality and expensive computations to get qualified data. These challenges make the uncertainty of this topic higher.

The fourth trend about estimating solar energy potential on house roofs virtually has a low uncertainty, since for example, Germany has introduced a law that makes PV panels compulsory on new buildings. The impact is medium. The exploitation of the full PV roof potential could fundamentally change the urban landscape. In addition to the generation of green energy, the associated positive side effects, such as building cooling, are a driving argument for the use of PV systems on roofs. However, there are still some barriers that need to be overcome, such as the complex processes involved in pattern recognition and the availability of high-resolution satellite imagery.

References

[1] Bhaeadwaj Anshu, Pratah Jain, and Ghosh Saptak. "FUTURE OF SOLAR PHOTOVOLTAIC Deployment, investment, technology, grid integration and socio-economic aspects". In: (2019).

[2] Imane Sebari Yahya Zefri Achraf ElKettani and Sara Ait Lamallam. "Thermal Infrared and Visual Inspection of Photovoltaic Installations by UAV Photogrammetry—Application Case: Morocco". In: (2018).

[3] Jonathan Kurén et al. *Fault detection AI for Solar Panels*. uppsala universitet, 2020.

[4] Shaveta Arora Nitu Rana. *A Review on Surface Defect Detection of Solar Cells Using Machine Learning*. Springer, 2020. Chap. 29.

[5] Dong Ho Lee and Jong Hwa Park. "Developing Inspection Methodology of Solar Energy Plants by Thermal Infrared Sensor on Board Unmanned Aerial Vehicles". In: (2019).

[6] Matteo Moncecchi, Stefano Meneghello, and Marco Merlo. "Energy Sharing in Renewable Energy Communities: the Italian Case". In: *2020 55th International Universities Power Engineering Conference (UPEC)*. 2020, pp. 1–6. DOI: `10.1109/UPEC49904.2020.9209813`.

[7] J.R Wijesingha et al. "Smart Residential Energy Management System (REMS) Using Machine Learning". In: *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*. 2021, pp. 90–95. DOI: `10.1109/ICCIKE51210.2021.9410779`.

[8] Gonzague Henri and Ning Lu. "A Supervised Machine Learning Approach to Control Energy Storage Devices". In: *2020 IEEE Power Energy Society General Meeting (PESGM)*. 2020, pp. 1–1. DOI: `10.1109/PESGM41954.2020.9281748`.

[9] Jongwoo Choi, Youngmee Shin, and Il-Woo Lee. "An Energy Sharing Framework for Individuals in the Smart Community". In: *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. 2016, pp. 1390–1391. DOI: `10.1109/HPCC-SmartCity-DSS.2016.0197`.

[10] Karry Kailin Hsu. "AI city's solar energy collaborative commerce and sharing economy". In: *Proceedings of the International Conference on Electronic Business (ICEB)*. 2018, pp. 98–105.

[11] Rodrigo Henriquez-Auba et al. "Sharing economy and optimal investment decisions for distributed solar generation". In: *Applied Energy* 294 (2021), p. 117029.

[12] François Lafond et al. "How well do experience curves predict technological progress? A method for making distributional forecasts". In: *Technological Forecasting and Social Change* 128 (2018), pp. 104–117.

[13] Asif Mahmood and Jin-Liang Wang. "Machine learning for high performance organic solar cells: current scenario and future prospects". In: *Energy Environ. Sci.* 14 (1 2021), pp. 90–105. DOI: `10.1039/D0EE02838J`. URL: `http://dx.doi.org/10.1039/D0EE02838J`.

[14] Ming-Yue Sui et al. "Nonfullerene Acceptors for Organic Photovoltaics: From Conformation Effect to Power Conversion Efficiencies Prediction". In: *Solar RRL* 3.11 (2019), p. 1900258.

[15] Zhi-Wen Zhao et al. "Computational Identification of Novel Families of Nonfullerene Acceptors by Modification of Known Compounds". In: *The journal of physical chemistry letters* 12 (2021), pp. 5009–5015.

[16] Harikrishna Sahu and Haibo Ma. "Unraveling correlations between molecular properties and device parameters of organic solar cells using machine learning". In: *The journal of physical chemistry letters* 10.22 (2019), pp. 7277–7284.

[17] Asif Mahmood and Jin-Liang Wang. "Machine learning for high performance organic solar cells: current scenario and future prospects". In: *Energy & Environmental Science* 14.1 (2021), pp. 90–105.

[18] Harshita Chopra. *Machine Learning For Roof Detection and Solar Panel Assessment.* 2020. URL: https://medium.com/omdena/machine-learning-for-roof-detection-and-solar-panel-installment-9ddeb0b70109 (visited on 09/02/2021).

[19] Pieter Gagnon et al. *Rooftop solar photovoltaic technical potential in the United States. A detailed assessment.* Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2016.

[20] Nour Daouk. *Sunny spells: How SunPower puts solar on your roof with AI Platform.* 2019. URL: https://cloud.google.com/blog/products/ai-machine-learning/how-sunpower-puts-solar-on-your-roof-with-ai-platform (visited on 09/02/2021).

[21] Acar Ahmet, Ceren Sarı Ayşe, and Taranto Yael. "Rooftop solar energy potential in buildings – financing models and policies for the deployment of rooftop solar energy systems in Turkey". In: (2020).

[22] Gaëtan Masson et al. "A snapshot of global PV markets-the latest survey results on PV markets and policies from the IEA PVPS Programme in 2017". In: *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)(A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC).* IEEE. 2018, pp. 3825–3828.

[23] Riel Martina. "Klimaschutzgesetz § 8a und § 8b: Die neue Photovoltaik-Pflicht in Baden-Württemberg". In: (2021).

[24] Kasun Perera, Zeyar Aung, and Wei Woon. "Machine Learning Techniques for Supporting Renewable Energy Generation and Integration: A Survey". In: vol. 8817. Sept. 2014, pp. 81–96. ISBN: 978-3-319-13289-1. DOI: 10.1007/978-3-319-13290-7_7.

[25] Akash Kumar and Indu Sreedevi. "Solar Potential Analysis of Rooftops Using Satellite Imagery". In: (Dec. 2018).

[26] Yash Kotak et al. "Installation of roof-top solar PV modules and their impact on building cooling load". In: *Building Services Engineering Research and Technology* 35 (Mar. 2014). DOI: 10.1177/0143624414527098.

Chapter 4

Artificial Intelligence for Smart Cities

Bali, Adrian
Garidis, Leon
Gu, Yujia
Gulmez, Ahmet
Lin, Chieh
Michl, Adrian
Miskin, Prayag
Nguyen, Noah Binh
Smajli, Fatbardh
Wang, Zhe

Abstract

The rate of urbanization is surging, and the population centers are becoming denser. This comes with multiple challenges, ranging from urban planning to air pollution. Throughout this chapter, some crucial areas of cities will be analyzed to find possibilities to tackle problems relating to climate change in big cities with machine learning (ML). A sustainable urban model is necessary to regulate and control the consumption of resources to keep the environment intact. This can be achieved by changing transportation for the masses, by creating better public transport and limiting time in a car. A waste management system, which can monitor, collect and recycle all of the trash created, is an essential service of a smart city. The energy consumption of megacities needs to be reduced and produced from sustainable sources. All this needs to happen while creating a healthy environment for the population. ML can offer many promising solutions to the arising problems.

4.1 Introduction

The world's population is predicted to increase to 10.9 billion by 2100. This increase is especially taking place in urban areas, which are continuing to grow [1]. With the rise in city population, many challenges arise. Challenges that lie in increased transportation frequency, large amounts of waste generation, decreased air quality and an increased energy usage. Not to mention the overall complex urban planning that comes with the high population density.

To efficiently handle the meteoric growth in urbanization and its effects, many countries have proposed the concept of smart cities to effectively manage the resources and optimize energy consumption.

Machine Learning techniques are known for their ability to handle large sets of messy, error-prone data. They use algorithms that exploit the availability of unlabeled and labeled data to provide efficient resource management and personalized services in smart cities.

This report will focus on potential future trends, where machine learning and data science techniques are applied to smart city applications. Each trend includes relevant facts, corresponding key drivers, the resulting challenges that come along with them and are ultimately assessed in terms of their impact on the environment. At the end, the identified trends will be rated in a driver matrix by their impact and uncertainty, creating a comprehensive overview of the various applications.

4.2 Trends

4.2.1 Smart Urban Planning

Nowadays, to combat the adverse effects of global warming, the large and small districts are proposing a new city model, called "the smart city", representing a community of average technology size, interconnected and sustainable, comfortable, attractive, and secure. The cities consume 75% of worldwide energy production and generate 80% of CO_2 emissions. [2] Thus, a sustainable urban model is necessary to regulate and control the consumption of resources to keep the environment in check. The notion of a smart city is established by combining the knowledge society and the digital city. It is defined as a "multi-layer territorial system of innovation" made up of digital networks, individual intellectual capital, and the social capital of the city, which together constitute collective intelligence. [3]

Facts

- More than 50% of the global population is now urbanized (United Nations Report 2012) . The complexity of the social ecosystem in cities and urban areas has increased, making sustainability an important factor [3].

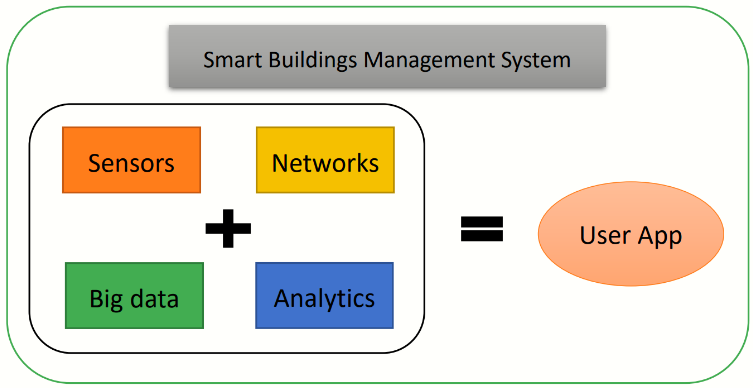

Figure 4.1: Component elements of smart buildings management system

- A lot of social-technical activities have been involved in digital transformation processes, and one purpose of that is to propose solutions for the sustainable smart city and the collaborative working between different parts in the whole city. Currently, smart building management systems require a variety of sensors, actuators and correspondingly the networks [4]. Figure 4.1 shows the component elements which define smart buildings management system. The goal is to monitor the situation of specific areas and improve a comfortable environment by using appropriate rules while saving energy.

- Uniform metadata for modeling smart buildings has been described by a lot of research recently. At this moment, speaking of smart buildings, an important concept: the Internet of Things (IoT) has to be mentioned, which is based on artificial intelligence (AI) and machine learning (ML) [5]. Several algorithms can be used in the smart building field, including conventional and statistical learning methods, probabilistic graphical modeling, neural network-based methods, and data mining algorithms.

Key Drivers

- For urban monitoring, modeling, and infrastructural analysis of cities, remote sensing is an essential tool.

- Detecting deterioration and irregularities with more available aerial and geo-tagged data of a city enable to provide the maintenance and sustainability of the city.

- Deep adversarial and representation learning techniques and research facilitate more autonomous urban planning by considering many aspects and features of a city. [6]

- Today, airports have become multifunctional conglomerates that generate substantial commercial development and a significant revenue stream on a micro and macro level. This gives way to a much broader concept where cities are emerging around airports, i.e., city airports are transforming into airport cities.

- AI has become the critical technological element in intelligent buildings; any modern construction designed to be intelligent must necessarily integrate connected objects. It is, therefore, essential to make the integration of the ML as dynamic as possible to allow the most enjoyable experience possible for its occupants [4]. In the meantime, the damage to the environment should be reduced as much as possible. Furthermore, it is impossible to make these buildings intelligent and dynamic without analyzing the data generated by this mass of connected objects.

- A large amount of data is created every second in the intelligent building context, so some solutions are proposed to process extensive scale data, such as Spark and Hadoop [7]. Both are large data frames, but they are not the same; we use them in different situations. If the operation and reporting requirements are static, we can choose MapReduce's operating mode. But if the data are not static, they need to be analyzed in streaming, like processing sensor data in a smart building, or if applications require a continuous operation, we need to use Spark. Lots of machine learning algorithms, which require several operations, are required for this case. So there are different clustering algorithms such as hierarchical grouping, K-means, self-organizing cards, and expectation maximizing grouping algorithm.

Challenges

- Generally, the sensor and hardware constraints might not let a very high spatial and spectral resolution of satellite images, limiting the accuracy in predictions and models.

- The rapid changes in cities and informality in certain regions make it difficult to understand and plan cities.

- The obtained datasets are primarily not very structured and labeled to gain an understanding of models.

- With the interaction between cyber and physical entities, the security vulnerabilities will increase as well, so that the feasibility of smart buildings will also decline. Therefore one challenge for intelligent management systems is to allow the generation of relevant characteristics, topological visualization for the security of a building, and verification of access control policies to meet security obligations [8].

Impact

Although urbanization is a long and tedious process until we can adequately achieve the goals of self-sustenance, we can see in practice the sound effects through some examples how the correct implementation of certain aspects can already lead to the betterment of the quality of life while reducing the impact on climate.

The impact of smart urban planning by retrofitting existing infrastructure to make it more eco-friendly and convenient can be most notably seen with Singapore, which is at the forefront of smart development with its Smart Nation vision. Examples of some of the initiatives within the framework of this project include:

- An open, cloud-based smart mobility platform offering data-driven shuttle bus service for commuters,

- A mobile application that lets users pay for short-term parking charges through their mobile devices at all existing coupon-based public car parks,

- An app developed by the Singapore Civil Defence Force(SCDF) in collaboration with the government acts to crowdsource for lifesavers. Whenever the SCDF is notified of an emergency, it sends a message to community first responders registered on the app.

- It makes the global vision of smart buildings possible that we have the ability to analyze huge amount of data based on many ML algorithms so that that building management can be simplified, energy consumption can be reduced, property and people can be kept safe, and an optimized living environment can be provided.

Another example of smart city planning can be seen in Norway, with its vision of the Oslo Airport City. Oslo Airport City (OAC) is an urban property development project which aims to build approximately 1 million square meters of a self-sustainable city planned from the ground up. It is designed to be powered only with renewable energy, with the excess being sold back into the grid. Sensor-based systems will operate automatic street and building lighting along with waste management and security. Only electric vehicles will be permitted, with planners eventually phasing everything towards self-driving vehicles.

4.2.2 Transportation

With more people expected to migrate to cities in the following decades, we believe consumers will begin to adopt a more digital way of life and a more connected and efficient car utilization. During the forecast period, advancements in IoT (internet of things) and other technical breakthroughs are projected to boost smart cities' transportation. Soon, rapid advances in information and communications, such as cloud, mobility, data communication, and sensors, are expected to drive the transportation for smart cities market. One prominent example is

Car-to-X-Communication, which is closely interwoven with autonomous driving. Networked cars can connect and communicate in real-time with each other (car-to-car) and objects in their environment (for example, with mobile devices or networks: car-to-mobile). An efficient car utilization is recognizable in car-sharing. With more shared cars, people might refrain from buying a car, which can eventually decrease the number of vehicles in a city and reduce vehicle pollution. Vehicle pollution contributes to global warming as greenhouse gases heat the planet and deplete the ozone layer. This is causing the average global temperatures to rise and finally leads to rising sea levels. Hybrid cars, electric cars, and alternative fuels will continue to help, but the sheer number of people and vehicles on the roads offsets these improvements. The rise in the number of vehicles on the roads will result in traffic jams and different kinds of pollution. Another factor of traffic congestion is the search for parking spaces. Nowadays, the search for parking spots in big cities like Munich is very time-consuming. One approach is an efficient routing system, but unfortunately, you need to know where free parking spaces are in advance. Therefore, finding that free parking spot is more efficient, using smart cameras and machine learning.

Facts

- Every car has an optimal speed range that results in minimum fuel consumption. When autos run at low speed, such as in a traffic jam and park, the fuel consumption is almost twice as the optimal speed.
- Under idling conditions, due to incomplete combustion in the cylinder, there are a lot of harmful and toxic components, i.e., hydrocarbons, carbon monoxide, and nitrogen oxides, which means traffic jams can pollute the air.
- According to [9], $30 - 45\%$ of traffic in urban areas is comprised of vehicles searching for parking lots.
- Taxis and shared cars are not used rationally, and there is a lot of waste of resources. For example, two people ride two cars separately, and the routes of the two cars overlap to a certain extent. This situation wastes fuel and increases vehicle exhaust emissions.

Key Drivers

- Further improvement of technologies like sensors, cloud, computational power is the cornerstone for the transport sector in smart cities.
- Route recommendation system: use Knowledge Graph Representation method to forecast the traffic speed or traffic flow in the future to plan a route that with less congestion [10].
- Ride-sharing system: Users send a request (including the earliest start time, the origin, destination, etc.) through smartphones to a cloud. The

cloud can arrange for the users to travel together based on route and time [11].

- Smart parking system: detect occupied and free parking spaces using cameras, computer vision, and machine learning.

- Autonomous driving: Deep reinforcement learning trains an agent to learn how to act under different scenarios. Autonomous driving can reduce the amount of braking which indirectly improves fuel economy [12].

Challenges

- For the route recommendation systems, more external factors should be considered in the knowledge graph [13]. For example, festivals may make some roads congested, which should be considered as an influenced factor.

- For Ride-sharing systems, more preferences should be concerned, e.g., the maximum number of shared persons, the pick-up/drop-off point selection, and the point of interest selection [11].

- Parking space detection is either very hardware expensive, lacks detail, or does not scale well for industrial applications.

- Use of smart cameras in parking garages and robustness against extreme weather conditions.

- Most methods are trained based on data sets, but real-time data should be collected to feed the model in real applications. This requires fast computation speed of the algorithm and hardware.

Impact

By linking to a potential mobile app or device, users get instant access to communication channels and data that can help them avoid traffic jams, report an accident, find an available rental or micromobile vehicle, or even find a parking spot. Smart cities and the Internet of Things will eventually work in tandem with vehicle technologies to make these areas safer, more sustainable, streamlined, and efficient. The route recommendation system can reduce congestion, consequently decreasing the emissions of greenhouse gases. The smart parking system can help the driver find free parking space conveniently within a shorter time, which will result in less traffic congestion. A Ride-sharing system improves energy efficiency by taking more passengers on a single journey.

4.2.3 Waste Management

With rising populations in cities, the amount of waste generated increases. This poses multiple challenges. First, the sheer amount of waste needs to be collected and sorted. The non-recyclable waste will be required to be disposed of environmentally friendly, while the recyclable trash needs to be sorted further and reused. Secondly, recycling's importance is constantly growing. This has two main reasons: there are not enough facilities and areas to dispose of all the trash and waste contains valuable materials that should not get wasted. Here, ML can assist in identifying areas of large pollution, sorting large amounts of trash for recycling, and optimizing the collection routes of the garbage trucks.

But other resources are too valuable to be wasteful with; water is increasingly becoming a scarce resource. Water management, especially excess water usage, should be controlled further. With the help of AI, it will be possible to quickly identify and stop water losses through leaks and broken pipes.

Facts

- In 2016, it is estimated that 2.01 billion tonnes of municipal solid waste were generated. If nothing changes, this number is expected to grow to 3.40 billion tonnes by 2050 [14].

- Overall, 37% of trash ends up in some landfill, while 33% is openly dumped and only 13.5% of trash is recycled [14].

- Waste generation is directly correlated with urbanization. High-income countries and economies are more urbanized, and they generate more waste per capita and in total [14].

- Landfill gas is a natural byproduct of the decomposition of organic material in landfills. It is composed of roughly 50% methane, 50% carbon dioxide (CO_2), and a small amount of non-methane organic compounds. Methane is a potent greenhouse gas 28 to 36 times more effective than CO_2 at trapping heat in the atmosphere over a 100-year period [14].

Key Drivers

- *Greyparrot* is a British company that designs automated waste composition analysis systems. Their system combines visible image data with live image processing and analysis, using AI deep learning techniques to recognize and distinguish waste types. An example of their algorithm at work can be seen in figure **??**. This approach is faster, more effective, and more accurate than manual spot sampling. With the more accurate database, municipalities can then optimize their recycling facilities [15].

- Today, less than 5% of rare earth are recycled from end-of-life devices [16]. This is partly due to today's incredibly complex products and the difficulty to extract the specific metals [17]. These materials are scattered across millions of devices, objects, and buildings, making it extremely difficult to amass enough material to recycle them.

- A lot of other raw materials are starting to become rarer. Sand, the most-consumed natural resource after water, is becoming harder to find [18].
- The rivers in cities are a confluence of trash. Being able to monitor, analyze and remove the trash from the river allows recycling material and keeping the environment clean. *The Ocean Cleanup* uses ML to monitor plastic pollution in rivers [19]. This can help in reducing openly dumped trash and increase the recycling rate.

Challenges

- In order to be able to manage all the produced waste, every source needs to be known, monitored, and able to extract the waste. Due to the many possible sources of waste, ranging from sewage, leaks, and openly dumped trash to stored electronics in drawers, it can be difficult to figure out where the materials are coming from and their composition.
- Just minimizing and collecting all the waste is already a big step towards good waste management. Yet, something needs to be done with all the material. It either needs to be appropriately discarded or recycled.
- It is essential that the second life of recycled material is put to good use. There is no point in going through the entire sorting and recycling process, just for it to be made into a new useless item, which will end up in the trash again. This would create an unnecessary and endless process of recycling.
- Another critical but challenging area is construction materials. The sheer amount of concrete and steel needed is extremely unsustainable in production, and they are rarely reused. ML could minimize the amount of concrete in certain areas or material science with materials such as gradient concrete.
- Overfilled waste bins eventually lead to incorrect waste disposal or waste ending up in the environment. Knowing where waste collection is needed and which route the garbage collectors should drive, not only to get there as soon as possible but also to reduce emission by avoiding unnecessary driving routes, is challenging. Therefore, a precise forecast of waste generation and a smart waste collection system is necessary to avoid overfilled waste bins and provide an efficient waste-collecting procedure. ML techniques are proven to be efficient in forecasting and solving traveling salesman problems to maximize the utilization of resources and minimize the cost of each waste collection route.

Impact

The overall objective of a climate-friendly circular economy could have an enormous impact on the environment. In the future, the significance of waste management will increase, not only through financial incentives but through pure necessity. The development of a process to reuse rare materials is of paramount importance. But also the construction and household waste need to be reduced. The high consumption lifestyle from people in urban regions is not sustainable, if their waste products cannot be managed and reused. Especially optimizing recycling procedures of all sorts of materials is highly valuable in fighting for a better climate.

4.2.4 Energy Efficiency

A large part of a city's CO_2 emissions is caused by its energy supply. For this reason, it is natural to try to use energy resources as efficiently as possible. New machine learning can make a considerable contribution here and open up new application possibilities. In general, a distinction is made between the city's energy supply and the individual consumers in the household. IoT-based approaches can help to improve the storage and management of energy reserves by estimating future energy demand and thus compensating for energy peaks that have so far been covered by conventional power plants [20] [21].

Much of the electrical energy that a municipality has to put up is due to urban lighting. ML approaches can help improve the efficiency of street lighting [22]. And beyond that, promising approaches are being explored on how to implement a common heating system to heat entire neighborhoods together via, for example, hot water, thus significantly improving efficiency [23].

Another area of application is offered by IoT-based approaches for consumers individually to improve energy efficiency. Already today, flats or houses are equipped with smart-home capable devices that, for example, automatically switch lights on and off or automatically close windows to save energy [24]. Collecting and analyzing energy data from individual homes will provide better insights into energy consumption and increase efficiency. ML approaches are ideally suited here as they are particularly good at dealing with this large amount of data and will continue to be one of the biggest drivers for improvement in the field, in which a wide range of international companies are already active [25].

Facts

- Heating, cooling, and domestic water make up more than 50% of building energy consumption [26]. Consequently, reducing those aspects is crucial to building energy saving.

- Over the years, better building standards and, therefore, better insulation led to smaller energy consumption in buildings. Out of the relevant sectors, warm water demand showed the slightest improvement and even grew the most in proportion; hence it's worth considering.

- IoT-based approaches can help to estimate energy demand and increase efficiency.

- Already today, a large number of international companies are active in each area, which underlines the outstanding importance and the great potential.

Key Drivers

- The application of ML methods in buildings has focused on heating and cooling demand. Hence, different models have been implemented successfully. [26] [27].
- The energy demand of smart lighting can be highly reduced with today's LED technology.
- Domestic hot water demand is less explored so far. The reliable prediction would enable the use of more efficient water heating systems. Simple ML models showed some success [28].
- With the help of ML, home appliances that make use of IoT can be used for energy monitoring, forecasting the energy consumption of buildings, and enable automated decisions [29].

Challenges

- Human behavior is highly stochastic and therefore hard to predict. In case of hot water demand, some regularities could be observed and exploited by the models, and savings were possible. However, new models and more diverse data are needed for more reliable results.
- Expanding and improving existing urban infrastructures is expensive or partly not possible.
- The possibility of a direct, as well as, indirect rebound effect is very present. ML approaches and IoT devices consume energy themselves. If the energy saved is less than the electricity consumed for the calculations, we speak of a direct rebound effect. Likewise, one should not underestimate the possibility of an indirect rebound. Improving efficiency will drive down the price. Individuals tend to spend the money saved on other things like travel, which would negate the energy savings.

Impact

The impacts are clear and obvious, as they are precisely aimed at reducing energy consumption and thus saving greenhouse gases. Apart from energy cost savings, a reliable prediction would also mean lower maintenance costs and more equipment life. Also, the energy supply of a city is an order of magnitude that can have a direct, measurable impact. This is why the potential is so outstandingly large and is hotly debated in both research and business. Nevertheless, the complexity of the problem is enormous. Converting or expanding already old urban structures is expensive or, in some cases, not even possible. The past has shown, that efficiency improvements often end up in rebound effects and have the opposite effect.

4.2.5 Air Quality

The air quality of a city may be considered a critical element in establishing a pollution index and how effectively the city's industry and population are governed. Air pollution has been a severe issue for the public and governments. Moreover, air pollution has a significant impact on the ecosystem and human health, resulting in acid rain, global warming, heart disease, and skin cancer.

Air pollution is a mixture of particles and gases that may accumulate to harmful levels outdoors and indoors. It can have a variety of consequences, including increased illness risks and increasing temperatures. Pollutants such as soot, smoke, mold, pollen, methane, and carbon dioxide are a few examples.

The Air Quality Index (AQI), which evaluates air quality across the country based on concentrations of five primary pollutants: ground-level ozone, particle pollution (or particulate matter), carbon monoxide, sulfur dioxide, and nitrogen dioxide. This index is one measure of outdoor air pollution.

Facts

- According to research from the World Health Organization (WHO), more than nine out of 10 of the world's population (92%) live in places where air pollution exceeds safe limits [30].

- Every year, household air pollution causes around 3.8 million premature deaths, the great majority of which occur in developing countries, with women and children accounting for approximately 60% of those deaths [31][32].

- Air pollution is to account for 26% of deaths from ischemic heart disease, 24% of deaths from strokes, 43% of deaths from chronic obstructive pulmonary disease, and 29% of lung cancer deaths [33][32].

Key Drivers

- Glasgow has been adopting the Sensing the City pilot project to monitor air quality and reduce emissions more efficiently and cost-effectively. Sensing the City uses the Libelium IoT Sensor node to minimize pollution through mobile monitoring. This low-cost method supports Glasgow's high-cost static sensing stations [34].

- BreezoMeter combines big data and machine learning technology to provide relevant, real-time, location-based air quality and pollen data. Incorporating numerous data sources, such as government monitoring stations, real-time traffic information, meteorological conditions, and CAMS data, BreezoMeter shows with high accuracy what is in the air we breathe and can give personalized health recommendations. The procedure of BreezoMeter is starting with collecting data from various sources to validating the data to produce geographical data points. Moreover, machine learning techniques are applied to model the data and lastly providing the results for users [35][36].

- Real-time and location-based measurements of indoor or outdoor air quality will help citizens make better and more informed decisions about where they spend their time.

- Connected devices will be more prepared to assist in the cleaning of the air in people's immediate environments, and therefore, reducing people's exposure to harmful air pollutants.

- Adding sensors to existing infrastructure, for example connected sensors, is another opportunity to expand data collection. Even though they are still less accurate and less regulated than official air monitoring stations deployed by governments, connected sensors are smaller and less expensive. The data gathered from these more widely installed sensors will aid in the collecting of massive amounts of data. Furthermore, rather than serving as a replacement for big data and ML models, their data should be used as an additional data source [36].

Challenges

- Performance variations due to various data sources and real-time information are a big challenge. For example, information on more than 5 billion people living in 67 countries is provided via the BreezoMeter API in real-time, accurately within 300 meters. BreezoMeter enhances air monitoring data with information from a wide range of independent sources, including local weather and traffic conditions, air pollution dispersion models, and satellites. As each calculation generates 1.5 terabytes of data, BreezoMeter must overcome massive performance difficulties in order to avoid falling behind[37].

- The affordability and sufficiency of monitoring stations should be considered. At the moment, Barcelona has 11 air quality monitoring stations. For a city of over a hundred square kilometers, these 11 quality monitoring stations are not a lot. That equals to one sensor every ten square kilometers in Barcelona. These monitoring stations have several disadvantages: they take up a lot of space, they are expensive (both in terms of installation and maintenance), and they only give data from a single location in space (cannot capture the hyper-local variations) [38][39].

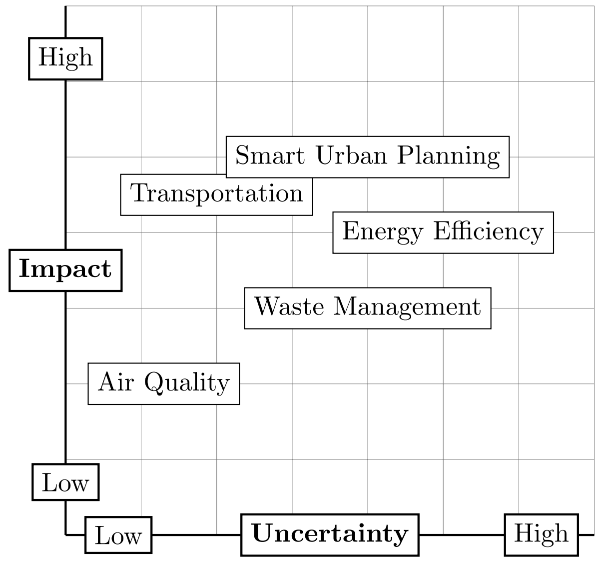

Figure 4.2: Driver matrix

Impact

Investigating air quality allows for a better understanding of the root causes of pollution and helps cities identify areas of opportunity to improve air quality. Therefore, tackling public health issues and determining solutions for a better climate. Having Air Quality Monitoring systems is a good first step, but the value lies in the use cases: How can this data be employed to make a genuine difference in air quality?

4.3 Conclusion

The increase in population, especially in dense urban environments, will cause many challenges in the future. The cities will have to grow and handle the bigger population, all while become more environmentally friendly. The discussed ML applications can help alleviate some of the challenges and create long lasting solutions. By using ML in the planning stages of infrastructure projects, suitable expansions can be build that will shape the cities for the future. In combination with transportation optimization, correct waste management and improved energy efficiencies across sectors, there is a increased possibility to better the environmental impact of cities. Not only will the air quality improve and therefore save people lives, but overall, the pressure on the environment will be decreased. The overall goal, of combating climate change in the most densely populated areas, can hopefully be achieved with the integration of ML in this aspect of the evolving human environment.

References

[1] Department of Economic and Social Affairs: Population Division. *World Population Prospects 2019*. United Nations, 2019.

[2] George Cristian Lazaroiu and Mariacristina Roscia. "Definition methodology for the smart cities model". In: *Energy* 47.1 (2012). Asia-Pacific Forum on Renewable Energy 2011, pp. 326–332. ISSN: 0360-5442. DOI: `https://doi.org/10.1016/j.energy.2012.09.028`.

[3] Sujata Joshi et al. "Developing Smart Cities: An Integrated Framework". In: *Procedia Computer Science* 93 (2016). Proceedings of the 6th International Conference on Advances in Computing and Communications, pp. 902–909. ISSN: 1877-0509. DOI: `https://doi.org/10.1016/j.procs.2016.07.258`.

[4] Abdellah Daissaoui et al. "IoT and Big Data Analytics for Smart Buildings: A Survey". In: *Procedia Computer Science* 170 (2020), pp. 161–168. ISSN: 18770509. DOI: `10.1016/j.procs.2020.03.021`.

[5] Biljana L. Risteska Stojkoska and Kire V. Trivodaliev. "A review of Internet of Things for smart home: Challenges and solutions". In: *Journal of Cleaner Production* 140 (2017), pp. 1454–1464. ISSN: 09596526. DOI: `10.1016/j.jclepro.2016.10.006`.

[6] Dongjie Wang et al. *Reimagining City Configuration: Automated Urban Planning via Adversarial Learning*. Vol. 1. 1. Association for Computing Machinery, 2020, pp. 497–506. ISBN: 9781450380195. DOI: `10.1145/3397536.3422268`. arXiv: `2008.09912`.

[7] Lamia Karim, Azedine Boulmakoul, and Ahmed Lbath. "Real time analytics of urban congestion trajectories on hadoop-mongoDB cloud ecosystem". In: *Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing*. Ed. by Hani Hamdan et al. New York, NY, USA: ACM, 3222017, pp. 1–11. ISBN: 9781450347747. DOI: `10.1145/3018896.3018923`.

[8] Liliana Pasquale et al. "Topology aware adaptive security". In: *Proceedings of the 9th International Symposium on Software Engineering for Adaptive and Self-Managing Systems - SEAMS 2014*. Ed. by Gregor Engels and Nelly Bencomo. New York, New York, USA: ACM Press, 2014, pp. 43–48. ISBN: 9781450328647. DOI: `10.1145/2593929.2593939`.

[9] Guo Chao Alex Peng, Miguel Baptista Nunes, and Luqing Zheng. *Impacts of low citizen awareness and usage in smart city services: the case of London's smart parking system*. Springer-Verlag Berlin Heidelberg 2016, 2016. URL: `https://www.researchgate.net/publication/310491800_Impacts_of_low_citizen_awareness_and_usage_in_smart_city_services_the_case_of_London's_smart_parking_system`.

[10] Y. Lin, Z. Liu, and M. Sun. ""Knowledge representation learning with entities, attributes and relations". In: *ethnicity* 1 (2016), pp. 41–52.

[11] Ming Zhu, Xiao-Yang Liu, and Xiaodong Wang. "An Online Ride-Sharing Path-Planning Strategy for Public Vehicle Systems". In: *IEEE Transactions on Intelligent Transportation Systems* 20.2 (2019), pp. 616–627. DOI: 10.1109/TITS.2018.2821003.

[12] Yao Deng et al. "Deep Learning-Based Autonomous Driving Systems: A Survey of Attacks and Defenses". In: *IEEE Transactions on Industrial Informatics* (2021), pp. 1–1. DOI: 10.1109/TII.2021.3071405.

[13] Jiawei Zhu et al. *KST-GCN: A Knowledge-Driven Spatial-Temporal Graph Convolutional Network for Traffic Forecasting*. 2020. arXiv: 2011.14992 [cs.LG].

[14] Silpa Kaza et al. *What a Waste 2.0 : A Global Snapshot of Solid Waste Management to 2050*. World Bank Publications, 2018. ISBN: 9781464813290. URL: https://openknowledge.worldbank.org/handle/10986/30317.

[15] Greyparrot AI Ltd. *AI-driven waste recognition system*. 2021. URL: https://www.greyparrot.ai/waste-composition-analysis-software (visited on 07/17/2021).

[16] recyclinginternational.com. *Is it now or never for rare earth recycling?* 2019. URL: https://recyclinginternational.com/non-ferrous-metals/rare-earth-metals/19629/ (visited on 08/20/2021).

[17] discovermagazine.com. *The World Is Running Out of Elements, and Researchers Are Looking in Unlikely Places for Replacements*. 2020. URL: https://www.discovermagazine.com/planet-earth/the-world-is-running-out-of-elements-and-researchers-are-looking-in-unlikely (visited on 08/20/2021).

[18] Vince Beiser. "Why the world is running out of sand". In: *BBC* (Nov. 18, 2019). URL: https://www.bbc.com/future/article/20191108-why-the-world-is-running-out-of-sand (visited on 08/20/2021).

[19] Colin van Lieshout et al. "Automated River Plastic Monitoring Using Deep Learning and Cameras". In: *Earth and Space Science* 7.8 (2020). e2019EA000960 10.1029/2019EA000960, e2019EA000960. DOI: https://doi.org/10.1029/2019EA000960. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019EA000960. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019EA000960.

[20] Hêriş Golpîra and Salah Bahramara. "Internet-of-things-based optimal smart city energy management considering shiftable loads and energy storage". In: *Journal of Cleaner Production* 264 (2020), p. 121620. ISSN: 0959-6526. DOI: https://doi.org/10.1016/j.jclepro.2020.121620. URL: https://www.sciencedirect.com/science/article/pii/S095965262031667X.

[21] João Pedro Gouveia, Júlia Seixas, and George Giannakidis. "Smart City Energy Planning: Integrating Data and Tools". In: *Proceedings of the 25th International Conference Companion on World Wide Web*. WWW '16 Companion. Montréal, Québec, Canada: International World Wide Web Conferences Steering Committee, 2016, pp. 345–350. ISBN: 9781450341448. DOI: 10.1145/2872518.2888617. URL: https://doi.org/10.1145/2872518.2888617.

[22] Eleonora Riva Sanseverino et al. "Smart city and public lighting". In: *2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*. 2015, pp. 665–670. DOI: 10.1109/EEEIC.2015.7165244.

[23] İsmail Ilhan, Mehmet Karaköse, and Mustafa Yavaş. "Design and Simulation of Intelligent Central Heating System for Smart Buildings in Smart City". In: *2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG)*. 2019, pp. 233–237. DOI: 10.1109/SGCF.2019.8782356.

[24] Dae-man Han and Jae-hyun Lim. "Smart home energy management system using IEEE 802.15.4 and zigbee". In: *IEEE Transactions on Consumer Electronics* 56.3 (2010), pp. 1403–1410. DOI: 10.1109/TCE.2010.5606276.

[25] Li Jiang, Da-You Liu, and Bo Yang. "Smart home research". In: *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826)*. Vol. 2. 2004, 659–663 vol.2. DOI: 10.1109/ICMLC.2004.1382266.

[26] Yabin Guo et al. "Machine learning-based thermal response time ahead energy demand prediction for building heating systems". In: *Applied Energy* 221 (2018), pp. 16–27. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2018.03.125. URL: https://www.sciencedirect.com/science/article/pii/S030626191830463X.

[27] Sanjiban Sekhar Roy, Reetika Roy, and Valentina E. Balas. "Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM". In: *Renewable and Sustainable Energy Reviews* 82 (2018), pp. 4256–4268. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2017.05.249. URL: https://www.sciencedirect.com/science/article/pii/S1364032117308961.

[28] Amirreza Heidari et al. "Adaptive hot water production based on Supervised Learning". In: *Sustainable Cities and Society* 66 (2021), p. 102625. ISSN: 2210-6707. DOI: https://doi.org/10.1016/j.scs.2020.102625. URL: https://www.sciencedirect.com/science/article/pii/S2210670720308428.

[29] Rozeha A. Rashid et al. "Machine Learning for Smart Energy Monitoring of Home Appliances Using IoT". In: *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*. 2019, pp. 66–71. DOI: 10.1109/ICUFN.2019.8806026.

[30] Hannah Ritchie and Max Roser. “Outdoor Air Pollution”. In: *Our World in Data* (2019).

[31] Fahad Ahmed et al. “Impact of household air pollution on human health: source identification and systematic management approach”. In: *SN Applied Sciences* 1.5 (2019), p. 418.

[32] Thomson Reuters Foundation. *10 facts about air pollution on World Environment Day*. URL: `https://news.trust.org/item/20190604234358-dr1my`.

[33] Aaron J Cohen et al. “Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015”. In: *The Lancet* 389.10082 (2017), pp. 1907–1918.

[34] Smart Cities World. *Sensing in the city*. 2016.

[35] Evangelos Kosmidis et al. “hackAIR: Towards raising awareness about air quality in Europe by developing a collective online platform”. In: *ISPRS International Journal of Geo-Information* 7.5 (2018), p. 187.

[36] BreezoMeter. *Accurate Air Quality, Pollen, and Active Fires Information*. URL: `https://www.breezometer.com/` (visited on 08/31/2021).

[37] Google Cloud. *BreezoMeter case STUDY — Google Cloud*. URL: `https://cloud.google.com/customers/breezometer`.

[38] Jordi Massagué et al. “2005–2017 ozone trends and potential benefits of local measures as deduced from air quality measurements in the north of the Barcelona metropolitan area”. In: *Atmospheric Chemistry and Physics* 19.11 (2019), pp. 7445–7465.

[39] Zigurat Global Institute of Technology. *Air quality monitoring and management for Smart Cities*. Dec. 2020. URL: `https://www.e-zigurat.com/blog/en/smart-air-quality-management-for-smart-cities/`.

Chapter 5

Machine Learning in the Transportation Sector

Eddouss, Wassim
Hamila, Firasv
Huang, Yuan
Khammari, Syrine
Li, Yinxian
Peng, Xiangyuan
Petroulas, Aimilios
Serin, Melike
Song, Donghao
Zhang, Zhihao

Abstract

In recent years, people have attached great awareness to climate change due to its growing negative effects on our day-to-day lives. Reducing greenhouse gas emissions is, therefore, one of the most popular research topics of today. One of the main reasons for the greenhouse effect is transportation. With the development of machine learning technology, the existing transportation will also usher in great changes, which helps to solve environmental problems, such as slowing down global warming. In the future, people will use faster, environmentally more friendly, and interconnected public transportation. The experts will continue to develop more efficient route planning algorithms to achieve resource savings. And the governments also pay more attention to traffic management and use big data to evaluate the congestion. Finally, a complete intelligent transportation system will be established in conjunction with the entire city planning. Based on the above development trends, we will achieve more reasonable resource utilization in the transportation field, gradually reduce greenhouse gas emissions, save non-renewable energy, and improve our travel quality. So

far, people have achieved great results. Of course, there are still some technical and management challenges for which we need to further seek for solutions.

5.1 Introduction

Greenhouse gas(GHG) emissions trap heat and warm the planet. Over the past 150 years, human activity has been responsible for almost all of the increase in GHG in the atmosphere. The main source of GHG emissions from human activity is the burning of fossil fuels for transport. With the increased globalization of the economy, there has been a significant increase in the global transportation of all kinds of goods. It increases the pressure to reduce GHG emissions in the transportation sector. Machine Learning can be a game-changer. In the last decade, the employment of machine learning in various fields has shown its strong ability to derive useful information from large amounts of data. Machine learning has huge potential to reduce GHG in the transportation sector. On the one hand, machine learning can indirectly help to reduce GHG emissions by analyzing large amounts of data to find the main causes of GHG emissions in the transportation sector. On the other hand, it can directly reduce emissions by optimizing existing transport systems. An excellent prediction model can also enable travelers to rationalize their travel patterns, departure and arrival times, and routes, saving unnecessary time wastage and improving the efficiency of work life. It also allows government services to understand and predict road conditions and make early predictions on possible road congestion and traffic accidents, saving society's burden and allocating social resources rationally. In the second chapter, we will introduce several applications of machine learning in the field of transportation, and their challenges and impacts. Four trends will be in detail discussed: Future public transportation systems, Route Planning, Intelligent Transportation Systems, and Traffic Management. In the last chapter, we will compare these four trends in terms of both uncertainty and impact, and present the results in the driver matrix.

5.2 Trends

5.2.1 Public Transportation System

Our vision for the future development trend of public transportation is: the urban public transportation system has a unified search operation and payment platform, to share and save energy, and the entire system is highly intelligent and electricity-based. From small to short-distance shared bicycles and skateboards to large-to-long-distance shared cars, we only need to enter the origin and destination in the system. The system will give you the simplest and fastest travel plan at the planning office, and then assist with the future digital currency as the basis, which completes the payment quickly. The core point is that shared cars are different from traditional buses and shared ride-hailing methods.

All future shared cars will be unmanned electric vehicles, and routes can be dynamically planned according to needs. To meet the convenience and comfort of people's travel, we use the knowledge of machine learning to design different shared car capacities and numbers for different cities. At the same time, data science is used to dynamically plan travel modes and routes, so that people can reach their destinations faster, shorten the overall driving route, save energy and reduce emissions.

Facts

- By recurrent neural network, the public traffic system can "diagnose" the traffic density in a certain time, or a certain area with accidents or congestion. Then it can adjust the timetable, station location, or traffic control signals. The "Minimum Number of Vehicle Strategy" program from MIT can reduce the number of Manhattan taxis in the United States by 30 percentage[1].
- Future transportation will be able to provide more dynamic carpool travel. Car2go from Daimler, DriveNow, and ReachNow from BMW, Marven from GM are all competing in the shared transportation market. Studies have shown that using shared transportation instead of traditional personal travel will reduce the number of vehicles by about 57 percentage and the total travel distance by 34 percentage [2].
- Personal services are more perfect and integrated. "Metropolitan" APP has been launched and used in Shanghai, which can be used to scan code for rides, recharge IC cards, and check the status of bus and subway information. And at the end of the journey, unified payment can be realized.
- Different public transportation is suitable for various areas. It will be more economical to adopt a mixed approach. People have realized this point and are committed to strengthening the ties between all kinds of public transportation. The time and cost of choosing a hybrid travel mode can be reduced to 70 percent[3].

Key Drivers

- Autonomous driving is something that has a decisive influence on this system, but it also has a high degree of uncertainty. Research on autonomous driving is also very hotly launched. But so far, there is less fully integrated autonomous driving system, and most of them have only developed into assisted driving. It will take at least a few years for autonomous driving technology to mature.
- Data storage and processing are also important in the future public transportation system. No matter it is route planning or user information, as long as it is integrated into a system, the database is quite large and needs to be updated at all times. There is unclear about whether the

company can obtain and process this data, and what role the government department should play.

Challenges

- The first is the technical challenge. Fully autonomous driving technology is not yet mature, and cannot be applied to the future public transportation system. In the face of such a huge system, research in the field of machine learning can quickly make plans and provide users with a convenient and fast integrated travel mode. Now people always spend too much time waiting for the car, and the comfort is very poor. Without machine learning such problems will certainly hard to be solved.

- Then there is the challenge of security. The security of huge data is a serious problem. From the route search to the final payment, a series of processes produce large amounts of data. Most of the public transportation is led and operated by the city government. If the public transportation system needs such a high degree of intelligence in the future, will the government be able to operate? If not, is it safe enough for the company to process this data?

- Finally, there is a conceptual challenge. We are trying to use the future urban public transportation to make people abandon their private cars, which can be freed from the time and energy spent on travel, thereby reducing emissions and protecting the environment. But whether people can accept such travel and payment methods is a challenge for many countries.

Impact

With the unified scheduling of public transportation and the more complete transportation networks, the public demand for private cars will be greatly reduced. Correspondingly, the demand of land for garages and parking lots in cities has also decreased. Cities will have more disposable land. The increase in public transportation can also effectively avoid traffic congestion and reduce daily travel time, which improves people's travel experience and mood. With the use of electricity and renewable resources as transportation energy, emissions of carbon dioxide and nitrogen oxide gases are gradually decreasing, which slows down the process of climate warming. Moreover, due to the introduction of technologies such as autonomous driving of public transportation, the road safety index can be guaranteed. And there will be fewer accidents on the road caused by human factors such as fatigue driving and drunk driving.

5.2.2 Route Planning

Route planning is used to decide the route to take from the source to the destination. In this report, we will only discuss the Route Planning problem in the transportation sector. Unlike the Path Planning or Motion Planning problem in the Robot field, the map or graph is not available for the robot.), the Route Planning problem in the transportation field is a pure Combinatorial Optimization Problem(COP). Many COP algorithms can be direct applied to solve the Route Planning problem. Hence, in this report, we will focus on the current research in COP.

According to the research history of the vehicle navigation system, vehicle path planning algorithms can be divided into static path planning algorithms and dynamic path algorithm. Static path planning is based on physical geographic information and traffic rules as constraints to seek the shortest path. Static path planning algorithms have become increasingly mature, relatively simple, but their application is not very meaningful for the actual traffic conditions. Dynamic path planning is based on static path planning, combined with real-time traffic information on the pre-planned optimal route to the destination for timely adjustment to get the optimal path. Different vehicle path planning problems can be derived by imposing different constraints and changing the optimization objective.

One of the classical approaches to solving combinatorial optimization is to model it as a mixed-accuracy programming problem and then solve it using the branch-and-bound method. This process recursively slices the solution space into search trees, during which boundary values are computed, and subtrees that do not contain optimal values are trimmed. The tricky part here is that the heuristic strategy in the extended search tree is very delicate. On the one hand, it is time-consuming to get a good heuristic, and on the other hand, the best heuristic is different for different problems or different stages of the same problem. Several methods were proposed to optimize this search process. For example, the 2013 paper "Dash: Dynamic approach for switching heuristics"[4] proposed the Dynamic Approach for Switching Heuristics (DASH) method, which dynamically switches subproblems during the search process and selects the most appropriate heuristic algorithm. This idea also provides an excellent opportunity to introduce different machine learning methods, which have the advantages of machine learning methods and the advantages of optimality guarantees of traditional methods, so there has been much-related work in the industry in recent years. [5][6][7][8]

Facts

With the rise of artificial intelligence in recent years, Deep Neural Network (DNN) has become the dominant approach to machine learning. In addition to the common areas of vision, speech, natural language processing, and recommendation, it has also been applied to the field of COP and has delivered some good results.

- One of the typical tasks of COP is to optimize resource utilization. For example, in the Traffic Engineering field, we aim to find an efficient method to plan and distribute flows in a traffic network. In [2] they present a novel order dispatch algorithm to provide a more efficient way to optimize resource utilization. This algorithm has been successfully deployed in the production system of Didi Chuxing, a vehicle for hire company.

- In the work from Google [**chipplacement**] they present a learning-based approach to optimize the chip layout.

- Recent works have shown that attention-based RL models outperform recurrent neural network-based methods on these problems in terms of both effectiveness and efficiency.[9]

Key Drivers

- Many combinatorial optimization problems, such as the Travelling Salesman Problem(TSP) or the Vehicle Routing Problem(VRP), are based on a graph structure [1], which can be easily modeled by the existing graph embedding or network embedding technique. In such a technique, the graph information is embedded in a continuous node representation. The latest development of graph neural networks (GNN) can be used in modeling a graph combinatorial problem due to its strong capabilities in information embedding and belief propagation of graph topology [10]. Use GNN models to build an end-to-end Deep Reinforcement Learning (DRL) framework to solve combinatorial optimization problems, in particular, TSP and VRP.[8]

- The disadvantages of deep reinforcement learning are its lack of problem generality and its inability to provide optimal solutions. In contrast, although general and guaranteed to be optimal, constraint programming does not make it easy to make branching decisions when searching the solution space. It is therefore essential to combine the two and thus overcome the limitations of both. A combinatorial optimization problem is first modeled using dynamic programming and then encoded into deep reinforcement learning and constraint programming to link these two types of methods.[5]

Challenges

- Deep learning models are suitable for big data and require much training to tune a usable algorithm, and even more training data is needed for complex scenarios. Furthermore, the cost of data collection and scenario simulation will become increasingly expensive.

- Deep learning algorithms are end-to-end (input data, output results) decision systems where the computational process cannot be interpreted, i.e. there is no transparency.

- The modeling of Expert Systems (based on independent knowledge bases, e.g. maps, traffic rules) takes too long. It is too expensive, the knowledge bases may have errors, and multiple rules may be contradictory, thus creating fragile systems. Therefore, this approach cannot be used alone to build decision algorithms for autonomous driving.

Impact

According to the statistics from the U.S. Environmental Protection Agency, 29 percent of 2019 greenhouse gas emissions are from the transportation sector, which is 6 percent more than that from the industry sector. The transportation sector generates the largest share of greenhouse gas emissions in America. Light-Duty Vehicles contribute the most(58 percent) to greenhouse gas emissions. The most direct way to reduce greenhouse gas emissions in the transportation sector is by optimizing the route. In particular, optimizing the route of light vehicles has a significant potential to reduce greenhouse gas emissions. Because almost all drivers of light vehicles use navigation software when planning their routes. And nowadays, all navigation software offers only the shortest or fastest route options, which means that the weights of the edges are either distance or time. If they can provide the greenest route option, this is likely to reduce greenhouse gas emissions significantly.

Intelligent Transportation Systems (ITS) represent a futuristic solution to reducing greenhouse gas (GHG) emissions caused by the transportation sector, which account for 29% of total GHG emissions worldwide. Addressing these systems has several advantages such as developing smart cities, reducing energy consumption, reducing congestion, and maximizing road safety. In intelligent transport systems, the cooperation between humans, infrastructure (traffic lights, cameras, etc.), and vehicles are established with the help of several communication models, which can be represented as a single model called Vehicle-to-everything (V2X). More precisely, V2X generalizes vehicle to infrastructure (V2I), vehicle to vehicle (V2V), vehicle to network (V2N), and vehicle to pedestrian (V2P). These models provide important information on congestion and fairness for drivers and pedestrians and thus improve safety and traffic and reduce emissions. In addition, Artificial intelligence and machine learning have a significant impact in developing these models, improving them, and especially increasing their efficiency and accuracy [3].

Facts

- A V2X form of communication called Cellular Vehicle-to-Everything (C-V2X) is adopted by many car manufacturers. This technology uses the performance of 5G broadband and combines it with 4G for wider cellular coverage to establish long-range network communication between the different elements of intelligent transport systems. It also uses wifi for direct short-range communication using all the different models (V2I, V2V, and V2P). This technology can be used in many areas of modern transportation solutions such as collision prevention, platooning, cooperative driving, and even autonomous driving [11].

- In some big cities, the time spent in the traffic to find a parking spot reaches up to 30% of the total traffic. [12] With the communication intelligent transportation systems provide between the environment and the vehicles, both cruising time and the parking spots can be reduced.

Key Drivers

- As a result of the wireless connection of vehicles, the infrastructure, and the network, the traffic information or accessory knowledge on the environment can be shared between the agents of the system. It can help detect and prevent possible dangerous situations, which otherwise would have been overseen. [13] With a fully implemented V2X model, a safer driving experience and accident-free transportation can be achieved.

- V2X systems can improve road utilization and balance the traffic load distribution. [14] This reduction in congestion has some significant outcomes. Firstly, it reduces the time spent in the traffic. Especially in the big cities, with more personal vehicles and therefore higher traffic density, this could make a change in people's lives. Secondly and more importantly, less time in traffic leads to better fuel economy. When the fuel efficiency is maximized, the GHG emission is optimized.

Challenges

- One of the biggest challenges of V2X technology is the problem of battery degradation[15]. Although recent studies have shown that the latter is no longer a technical challenge, it is still an unsolved problem. Electric battery life depends on local conditions, such as the quality of grid service and the vehicle model. In order to convince users to participate in V2X technology, it is crucial to offer them financial compensation in case the battery life is reduced.

- Implementing V2X technology requires some adjustments to the distribution network. These adjustments can be costly, making this technology not only time-consuming but also exclusive to wealthy countries that can afford it. Some experts believe that major changes in distribution grids can be avoided if smart solutions are used. Smart solutions can be, for example, using machine learning algorithms to predict unknown grid variables instead of placing large numbers of smart meters all over the grid.

- Traffic congestion monitoring remains one of the key difficulties of V2X implementations. For this purpose, sensors or GPS can be placed in all vehicles. Additional monitoring systems can be installed in strategic parts of the city. However, it means that this system will produce a large amount of data, some of which might be considered outside of data protection regulations. Because installing stable cameras all over the city and movable ones inside vehicles lead to "being watched" by the system all the time. To solve this problem, different sensors, such as radar sensors, can be prefered. But it might decrease the accuracy of the collected data.

- Smart parking also requires smart placement of parking spots. To receive information on the parking sports, especially in the popular places such as shopping malls and city centers, a new infrastructure must be added.

Impact

Intelligent Transportation Systems affect the transportation sector on several levels. The current impact maybe not be significant enough due to the lack of the utilisation of ITS in the current transportations systems. However, its influence is growing day by day since many cities and vehicle manufacturer companies are starting to adapt Intelligent Transportation Systems. ITS respond to the predicament of large-scale automobile traffic efficiently and sustainably providing road safety for drivers, passengers, and pedestrians. It conducts real-time actions to ensure smooth traffic that contains much less congestion and more safety by providing a variety of services such as traffic management, automatic roadside inspections, variable traffic light and speed limits, and emergency management systems [3]. ITS doesn't only affect safety and congestion but it also has an economic impact. Using ITS, transport will cost less money since these systems provide optimal routing and have ride-sharing options. These systems also reduce greenhouse gas emissions, help in constructing smart cities, and lead to a greener lifestyle. Intelligent Transportation Systems mostly uses electrical or hydrogen-powered vehicles instead of fossil fuels based vehicles. It also provides optimal routing and avoids congestion which reduces energy consumption and thus saves energy and reduces emissions.

High emissions and environmental pollution caused by traffic congestion have become a huge and heavy burden on society. Therefore, the prediction of urban road network traffic flow and the rapid and accurate evaluation of traffic congestion is of great significance to the study of urban traffic solutions. Traffic management is a paradigm proposing a route server capable of handling the traffic in a city and balancing traffic flows accounting present and future traffic congestion conditions [16]. Many of the existing standards and algorithms are mentioned in [17], as well as an evaluation algorithm capable of calculating and evaluating traffic congestion levels based on microscopic traffic flow characteristics.

Facts

- Traffic management surely is connected with Intelligent transportation systems **??**; ITS widely uses information technology, traffic engineering, and behavioral science to reveal the rules of urban traffic, calculate traffic flow in real-time, and try to guide vehicles and their surrounding vehicles to avoid traffic congestion. The rapid development and deployment of ITS in recent years have produced a lot of real-time traffic information [18].

- Accurate traffic information prediction is important for many applications in Advanced Traffic Information System (ATIS) and vehicle energy management. If traffic congestion information can be correctly predicted in ATIS, it can be used to reduce the uncertainty of future traffic state, improve traffic mobility, and provide drivers with reliable travel time estimates, expected delays, and optional routes to their destinations [17].

Key Drivers

- The technology mentioned in [16] has played a significant role in practical applications. The proposed solution was validated in large-scale scenarios, such as the city of Valencia, with an area of 77.43 km2, and where the complexity increases due to the numerous routes that the vehicles requested to centralized traffic manager. The proposal improved travel times by 5%, and the average travel speed was 5% higher compared to the reference traffic for the entire city. They also focused on areas facing high congestion levels, such as the Ruzafa neighborhood in the city of Valencia, where they evaluated our proposed solution when injecting additional vehicles into this area to check the limits of our traffic balancing algorithm. In particular, the proposed solution was able to improve travel times by 8% compared to the default traffic conditions even under very high loads. In addition, the average travel speed with our proposed solution was 16% higher, meaning that vehicles arrived faster to their destinations.

- As future work, conduct further experiments are planning to conduct to determine the gain achieved by the approach in terms of CO2 emissions and fuel consumption.

- Paper[17] focuses on how to apply data science technologies on vehicular networks data to present a prediction method for traffic congestion based on both real-time and predicted traffic data. Two evaluation frameworks are established, and existing methods are used to compare and evaluate the accuracy and efficiency of the presented method. The main reference value of congestion evaluation is the occupancy rate of roads at a specific time. Two comparison frameworks are used to verify the algorithm and compare it with many other algorithms. The road network data used in the comparison are, respectively, from Cologne, Germany, and Karamay, Xinjiang, China. Traffic data are derived from real data and simulated data generated by SUMO.

- Based on the experimental results, the results can be summarized as follows: the congestion evaluation method proposed in this paper can accurately express the congestion degree of road network based on real-time and predicted traffic data based on high efficiency and fast original intention. This conclusion is supported by the following facts:

    1) For a single road, the processing time of 3000 seconds' sampling data is less than 0.0005 seconds on average, and the processing time of real-

time processing and real-time rendering of more than 70000 sampled data is less than 6s;

2) Without training cost, the algorithm itself is based on real-time or predictive traffic data in a short period of time, without the need for a large number of data as training sets and a large amount of time for training;

3) Accuracy is higher than other algorithms compared;

4) It can support congestion calculation and real-time rendering for different cities, different road networks, different road structures, different traffic modes, and different sampling data types. It is robust and can support larger road network and sample data processing and rendering when the hardware is extended.

Challenges

- Traffic management enables data from a wide variety of sources without actual integration between them. In parallel, the emerging technology of the Internet of Things will enable data exchange to numerous everyday devices that will integrate different sources more sustainably. However, this causes other important challenges such as managing a large number of devices, identify devices with different addresses to control the information flow. Alongside the handling of various data sources comes the problem of handling the huge amount and large variety of data that a big city can produce in terms of public transport. Alongside the handling of various data sources comes the problem of handling the huge amount and large variety of data that a big city can produce in terms of public transport. Furthermore, providing uncorrelated data asynchronously is another challenge due to non-integration among different sources. So TMS needs to find ways to fuse, aggregate, and manage data in a way that deals with heterogeneous data sets in an until now decentralized system. [19]

- Present different information into a compact traffic representation. Find the correct correlations between some information and the traffic, get the knowledge of how important each parameter is in the formation of traffic.

- Centralized route management scheme may lead to overhead that limits its potential performance. Moreover, it may be very inefficient for a very large number of vehicles and not be done in an acceptable time. A solution could be to let each vehicle perform route allocation, but it would be difficult to give each vehicle a spherical view of the traffic condition and also to suggest routes that will not later lead to traffic congestion in new areas. For that reason, a proper trade-off between efficiency and complexity should be decided.

- Sharing so many data could raise doubts about privacy issues and ensuring privacy and security for everyone should be guaranteed from the start for such implementation to work. So one challenge is the encryption of shared data and maintaining communication and efficiency in computing alongside security techniques.

Impact

The vast majority of emissions in the transport sector are caused from light private cars and other vehicles, so it is exactly the target of traffic management techniques. As apporaches for a traggic management are yet in an experimental level, not real implementation have been made so that a emmisions reduction can be measured. However results show that such techniques can succed in reducing the travel time as an average, which is directly related to GHG emissions through cars. We consider based on existing work [16] that the knowledge of implementing a centralized route management exists, and implementations will start in the upcoming years. Yet there are important challenges to be dealt with but since the technology for solving these challenges already exists, we classified this trend as a low uncertainty one. On the other hand, because this high percentage of private vehicles, that cause this traffic congestion, are responsible for the biggest part in emissions in the transportation sector and such approaches can be integrated in every big urban area we think that the impact of this trend should be above the middle. Even if it causes a small change for every city, the fact that it can be used for billions of cars around the world makes the trend worth mentioning and worth for further research and implementation. Last but not least, the upcoming autonomous vehicles can help even more in such strategies, and they can be immediately integrated in such systems.

5.3 Conclusion

We can see from the above four trends that machine learning will play an extremely important role in transportation in the future, so as to achieve our goal, which is to reduce greenhouse gas emissions, prevent global warming, and protect our living environment. Some of these four trends still have great uncertainties. For example, the unmanned driving required by both Public Transportation System and Intelligent Transportation System is currently immature, and the application is risky, which is still very large for us. Challenges are worthy of our continued research on breakthroughs. As for routes like Route Planning and Traffic Management, we can use some of the existing artificial intelligence knowledge and mature models to combine the achievements of science and technology with environmental protection concepts to achieve energy conservation and emission reduction, which has an important impact on our human development. . After all, trends are just our guesses about the future based on our current level of knowledge. It takes time to test how the future develops. But just like our seminar, when people keep talking about using artificial intelligence machine learning to improve the environment, and the development of various fields to reduce greenhouse gas emissions, more and more of us will pay

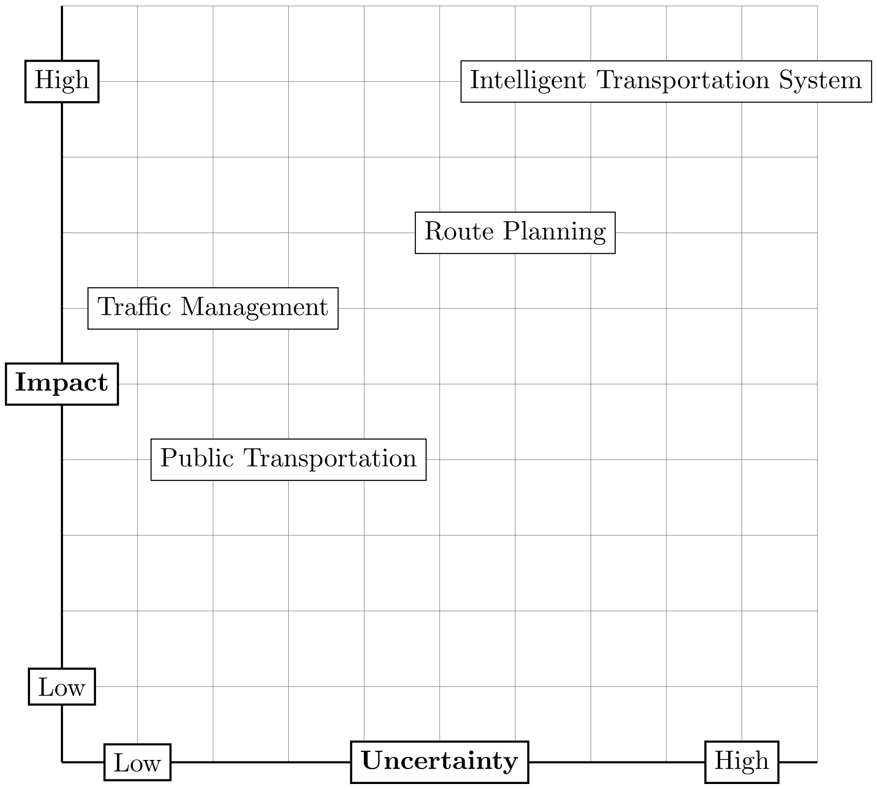

Figure 5.1: Driver matrix

more attention to science and technology and environmental protection. It is believed that our trend forecast will move in a positive direction.

References

[1] Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. “Machine learning for combinatorial optimization: a methodological tour d’horizon”. In: *European Journal of Operational Research* (2020).

[2] Zhe Xu et al. “Large-Scale Order Dispatch in On-Demand Ride-Hailing Platforms: A Learning and Planning Approach”. In: July 2018, pp. 905–913. DOI: 10.1145/3219819.3219824.

[3] Zineb Mahrez et al. “Smart Urban Mobility: When Mobility Systems Meet Smart Data”. In: *IEEE Transactions on Intelligent Transportation Systems* (2021).

[4] Giovanni Di Liberto et al. "Dash: Dynamic approach for switching heuristics". In: *European Journal of Operational Research* 248.3 (2016), pp. 943–953.

[5] Quentin Cappart et al. "Combining reinforcement learning and constraint programming for combinatorial optimization". In: *arXiv preprint arXiv:2006.01610* (2020).

[6] Maxime Gasse et al. "Exact combinatorial optimization with graph convolutional neural networks". In: *arXiv preprint arXiv:1906.01629* (2019).

[7] He He, Hal Daume III, and Jason M Eisner. "Learning to search in branch and bound algorithms". In: *Advances in neural information processing systems* 27 (2014), pp. 3293–3301.

[8] Kun Lei et al. "Solve routing problems with a residual edge-graph attention neural network". In: *arXiv preprint arXiv:2105.02730* (2021).

[9] Yunqiu Xu et al. "RLattention". In: *IEEE Transactions on Cybernetics* (2021).

[10] Mingxiang Chen et al. "Dynamic Partial Removal: A Neural Network Heuristic for Large Neighborhood Search". In: *arXiv preprint arXiv:2005.09330* (2020).

[11] GSMA Connecting Vehicles. "Today and in the 5G Era with C-V2X (Cellular Vehicle-to-Everything)". In: *GSMA: London, UK* (2019).

[12] Donald Shoup. "Cruising for parking". In: *Transport Policy* 13 (Feb. 2006), pp. 479–486. DOI: 10.1016/j.tranpol.2006.05.005.

[13] Shanzhi Chen et al. "Vehicle-to-Everything (v2x) Services Supported by LTE-Based Systems and 5G". In: *IEEE Communications Standards Magazine* 1.2 (2017), pp. 70–76. DOI: 10.1109/MCOMSTD.2017.1700015.

[14] Mate Boban et al. "Use cases, requirements, and design considerations for 5G V2X". In: *arXiv preprint arXiv:1712.01754* (2017).

[15] Christine Gschwendtner, Simon R. Sinsel, and Annegret Stephan. "Vehicle-to-X (V2X) implementation: An overview of predominate trial configurations and technical, social and regulatory challenges". In: *Renewable and Sustainable Energy Reviews* 145 (2021), p. 110977. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2021.110977. URL: https://www.sciencedirect.com/science/article/pii/S1364032121002690.

[16] Jorge Luis Zambrano-Martinez et al. "A Centralized Route-Management Solution for Autonomous Vehicles in Urban Areas". In: *Electronics* 8.7 (2019). ISSN: 2079-9292. URL: https://www.mdpi.com/2079-9292/8/7/722.

[17] Xu Yang et al. "Application of Data Science Technologies in Intelligent Prediction of Traffic Congestion". In: *Journal of Advanced Transportation* 2019 (Apr. 2019), p. 2915369. ISSN: 0197-6729. DOI: 10.1155/2019/2915369. URL: https://doi.org/10.1155/2019/2915369.

[18] A.D. Joseph et al. "Intelligent Transportation Systems". In: *IEEE Pervasive Computing* 5.4 (2006), pp. 63–67. DOI: 10.1109/MPRV.2006.77.

[19] Allan M de Souza et al. "Traffic management systems: A classification, review, challenges, and future perspectives". In: *International Journal of Distributed Sensor Networks* 13.4 (2017), p. 1550147716683612. DOI: 10.1177/1550147716683612. URL: https://doi.org/10.1177/1550147716683612.

Chapter 6

Artificial Intelligence to Predict Extreme Weathers

Ben Abdessamad, Mohamed Wael
Huang, Jiangnan
Hülder, Till
Zhang, Xin
Yu, Runyao
Xu, Qinyang
Smaoui, Khalil
Yilmaz, Sefa Ekin
Yu, Jiale
Zhong, Xuyang

Abstract

Extreme weathers like hurricanes, heat waves and their subsequent effects will cause huge damage on the infrastructures and human lives. Fortunately, the reliability of forecasting these weathers keeps improving thanks to the growing computational power, increasing number of observations and advanced physical models. Nowadays, artificial intelligence (AI) plays an important role in the accuracy improvement of the prediction results. Yet, how far the AI technology can lead us in predicting extreme weathers still remains a question. We intend to study the current progress of the implementation of AI technology in this area and its facing challenges. In this report, we select five possible trends. We show how AI is already used in the prediction of hail, temperature, wind, flood and segmentation of weather events. Among these the bias correction for extreme air temperature and precipitation prediction is considered to be the most promising. In all the trends we have selected, the main challenges of AI usages remain as the accuracy and amount of data. A larger range of effective

prediction both in time and space also worth pursuing.

6.1 Introduction

As a consequence of global warming, extreme weather events are much more common and destructive than before. The number and intensity of heat waves, major hurricanes and heavy downpours have increased in USA [1]. The cold wave and winter storm in the beginning of year 2021 have caused 20.4 billion dollars and 172 deaths in USA [2]. Moreover, it is expected in latest report that the extreme precipitation in Western Europe will occur more frequently, which has just caused severe flood in Germany in July, 2021 [3]. It is obvious that we should accept the fact that there is no run-away from the extreme weather. Accurate preparations such as a better prediction ability is needed urgently. Fortunately, the accuracy of the weather prediction has also gained progress in the past decades[4]. These achievements are mainly based on the improvements of observations, numerical models and data analysis. At the same time, the developing AI technology is promising in the study of environment[5]. This naturally leads to the investigation of AI in the prediction of extreme weathers. This report will list five possible trends of research directions in the following chapters, and each trend will be analyzed via its key drivers, challenges and possible impacts.

6.2 Trends

6.2.1 Machine Learning For Hail Prediction

Hail is an uncommon meteorological phenomenon in which small particles of ice fall from the sky, and the storms that produce hail reaching the ground are known as hailstorms. Hail is a solid form of rain made up of spheres or non-regular ice cubes, with a diameter of between 5 mm or 15 cm, and normally formed under some certain circumstances, for example the strong, upward motion of air freezing temperatures at lower heights. As a common, costly and potentially weather event, the hail is already a high-impact severe weather hazard, which can cause injuries to people, and damage to buildings, vehicles and crops. It is showed statistically, that hail annually causes in excess of 1 billion U.S. dollars of property damage and 1 billion of crop damage[6]. Another example is a single hailstorm during the afternoon rush hour in the Denver, Colorado, metropolitan area on 8 May 2017 resulted in 2.3 billion of insurance claims[7].

In order to reduce the economic impacts of severe hail, the need for accurate and timely detection, prediction and tracking are underscored, which allow individuals and businesses to take necessary action toward avoiding risk to their property and personal safety. However, accurate predictions of hail remain a challenge due to the rapid evolution of hail-producing convective storms, coupled with uncertainties and limitations of atmospheric observation data needed to properly resolve the small-scale convective environment.

Recently, the learning-based approaches have been proposed as powerful and efficient for hail predictions. We have reviewed multiple paper related to hailstorm using machine learning approaches, and summarize in this chapter the current situation, state of art methods, challenge and also an outlook of the future development with impacts. This subsection is assigned to Jiangnan Huang and Xuyang Zhong.

Facts

- The existing hail forecasting techniques, such as proximity sounding[8], numerical weather prediction model[9], have too many simplifications and therefore are sensitive to difference. Except that, it is often not calibrated and results comparatively less robustness.
- With the improvement of sensing and storing technologies, a large amount of weather data become available, and the data size will continue growing as radar imaging instruments continuously acquire data. The conventional approaches have significantly worse performance comparing to learning-based methods, when we want to process these gigantic data sets and predict qualitatively as well as quantitatively hailstorms.
- Many machine learning approaches have been proved in other scientific researches and practical applications that they obtain high accuracy, robustness and efficiency to solve such problem based on big data. Besides the computational efficiency, learning-based approaches don't require many prior information about model itself. Thus, they can also avoid too many assumptions, which leads the traditional approaches to complexity and worse performance.
- The learning-based approaches can take advantage of current meteorological data sets, such as HREFv2[10], for training, validation and test. For example, the input can be hailstorm size, spatial or temporal change and the output can be the probability of hail occurrence.

Key Drivers

- Many machine learning technologies have been used in the domain of hail research for detection, prediction, tracking and correction or calibration. In this section it will be introduced some main technical streams and evaluation methods.
- Regression
  - Elastic net regression: It's a specific penalized linear model regression to estimate shape and scale parameters of gamma distribution in [11].
  - Isotonic regression[12]: It's a model which calibrates the severe hail predictions toward the local storm reports and Storm Prediction Center's practically perfect output in [13].

  - Ridge regression with principle components analysis(PCA)[14]: The dimension of features is firstly reduced by PCA, and the PCA generated features are the input of ridge regression, which is used to predict the occurrence of hailstorm.

- Random Forest(RF): It's an ensemble learning method which trains multiple decision trees with randomly selected features for branching. It can be used for classification or regression and an example for hailstorm prediction can be found in [11].

- Neural network approaches

  - Autoencoder gives a better representation of the original input through lossy compression, but it carries out denoising and reconstructs the raw input. This input is then passed into a binary classifier, which predicts whether the output is a hailstorm or not[15].
  - One of the most popular learning-based solution is deep neural network(DNN) with convolution. Since the data set frequently are a large amount of radar images, the convolution operation is useful to extract features from them. DNN is a powerful approximation to real model and doesn't required much prior information. DNN is used for hailstorm prediction in [14], [16] and detection in [17].
  - Many more neural networks have been developed for hailstorm prediction, such as ResNet network[18], VGG16 network[19], nonlinear SVM[20], Bayesian neural network[21]. In [16], these approaches are also implemented as comparison.

- For evaluation there have been used usually statistics like confusion matrix, precision and recall in [16], BSS in [14] and graphical representation like ROC in [13].

Challenges

- The training data in [14] are simulated storm data, which were generated in convection allowing models(CAMs). They are often displaced in space, time, and intensity from the observed hailstorms, so the verification scores will generally be less skilled than the perfect model results.

- Data set limitation

  - Some no-hail images, which are used for training but contain hail-related features, may lead to miss-classification of model.
  - Some events are extremely localized, hail features are only in a small region on the image. It affects the performance of model.
  - If hailstorm and thunderstorm happen at the same time and same place, the case is often labelled as "Hail".

- The existing methods only carried out experiments on small data sets limited to a region, country, or location and a large number of input features, so the trained model might not work in other regions.
- Due to the data imbalance, Autoencoder and convolution neural network(CNN) will have a bad performance on minor data.

- DNN have have many possible configurations and parameter settings, it is time-consuming to find the best set of hyper-parameters.

Impact

Machine learning has shown its great potential to hailstorm prediction, but there are still rooms for improvement. For instance, CNN can be combined with RF. Since RF is more robust to data imbalance, and CNN has stronger learning capability, the combination may result in a better result. Besides, the actual researches lack the exploitation of temporal information, the introduction of recurrent neural network(RNN) may be helpful.

For hailstorm prediction, machine learning model is capable of giving satisfying results. It liberates technical staff from monitoring the weather information all day, and can give warning signal in real time. For example, as Figure 1 shown, the severe hailstorm spatial distributions can help technical staff and policy makers prevent disasters. However, for some developing countries, they do not have enough weather data to train a model. If some companies or institutions are willing to make their pre-trained models open-sourced, ML can better benefit humanity.

6.2.2 Bias Correction for Extreme Air Temperature and Precipitation Prediction

The Numerical Weather Prediction (NWP) model has been broadly utilized for forecasting air temperature and precipitation, yet by and large, it has a specific bias because of its coarse framework goal and absence of parameterizations. Revising the forecast bias of numerical climate prediction models is significant for serious climate alerts. Usually, the refined grid forecast requires direct correction on gridded forecast products, rather than amending forecast information just at individual climate stations.

AI is exceptionally useful in such cases, it can assist us with adjusting the bias that enables a better extreme climate forecast. In a recent research paper [22], various techniques were utilized for the bias correction such as random forest (RF), support vector regression (SVR), artificial neural network (ANN), and a multi-model ensemble (MME) to correct the Local Data Assimilation and Prediction System (LDAPS) which is a local NWP model over South Korea. It is a model that yields the next-day most extreme and least air temperatures in Seoul.

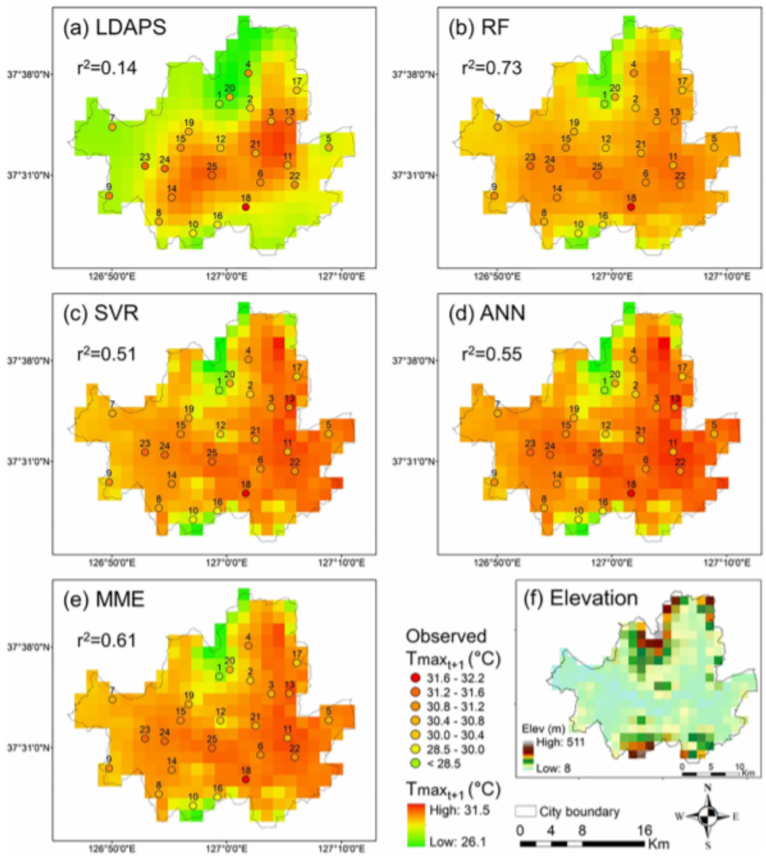

Figure 6.1: Map of spatial distribution of average forecasted $Tmax_{t+1}$ based on (a) LDAPS, (b) RF, (c) SVR, (d) ANN and (e) MME models, and (f) SRTM elevation aggregated to 1.5 km resolution [22].

Facts

- In the Republic of South Korea, 60 to 70% of the yearly precipitation sum [23] is generally produced throughout the summer time frame (July to September). In order to be ready for catastrophic events and guaranteeing the sustainability of water assets, efficient water management plans should be detailed for the weighty precipitation time frame. The determination of exact bias correction techniques can help in acquiring dependable projected precipitation changes across the Korean landmass for the late spring time frame later on, which can help in moderating expected dangers from catastrophic events.

- Forecasting $T_{max_{t+1}}$ [22] : The LDAPS model had an R^2 of 0.69, a bias of

$-0.85\ ^{\circ}C$, and an RMSE (one of the commonly used error-index statistics for observed and simulated data) of 2.08 $^{\circ}C$, whereas according to hindcast validation all bias correction models improved performance with R^2 ranging from 0.75 to 0.78, biases from -0.16 to -0.07 $^{\circ}C$ and RMSEs from 1.55 to 1.66 $^{\circ}C$. A comparison of the different used methods can be seen in figure

- Forecasting $T_{min_{t+1}}$ [22] : The LDAPS model had an R^2 of 0.77, a bias of 0.51 $^{\circ}C$, and an RMSE of 1.43 $^{\circ}C$, whereas the bias correction models showed R^2 from 0.86 to 0.87, biases from -0.03 to 0.03 $^{\circ}C$ and RMSEs from 0.98 to 1.02 $^{\circ}C$.

- A downside of climate-means-bias-correction is the requirement for a long preparing dataset [24], and since an estimate works best with a frozen model, the information base should be totally reconstructed at whatever point the model is updated, which requires a lot of computational assets. Therefore, routine improvements to the model are consolidated in the forecast-based products as soon as they are implemented.

Key Drivers

The most commonly used bias correction methods in the air temperature forecasting fields are the MOS and KF techniques.

The MOS improves forecasting accuracy by applying a statistical linear model developed between the past model results and observation data to the NWP model output. The Kalman Filter (KF) has been widely used to solve nonlinear problems. When forecasting air temperature, KF first bias-corrects NWP model output.

In addition to that, some machine learning approaches have been used to correct the bias of the NWP model's air temperature outputs. As referenced above various strategies of the machine learning approaches were utilized, for example, random forest (RF), support vector regression (SVR), artificial neural network (ANN), and a multi-model ensemble (MME). This Machine Learning techniques are not sensitive to the multi-collinearity of input variables, and thus can deal with many input variables.

- Random forest (RF): The first bias correction model, RF, is an ensemble machine learning algorithm that predicts a target variable from a set of predictors by growing multiple trees and aggregating their results. RF has been widely used to solve a multitude of classification and regression problems [22].

- Artificial neural network (ANN): ANN has an interconnected structure which emulates the operations and connectivity of biological neurons in human brain. This study used a multi-layer perceptron (MLP) neural network that consists of input, output and hidden layers with a back-propagation algorithm, which is the most popular due to its ease of training, meaning it has been widely used in many applications including forecasting [22].

⇒ ANN and RF are used to improve the minimum temperature forecasting skills of two NWP models. ECNWF and Local Area Model Italy (LAMI) in a region of the Italian Alps found that, compared to other approaches, RF yielded the best results with the advantage of an easily automated process.

- The SVR algorithm, aims to get the optimal hyperplane that fits the data and predicts with minimal empirical errors. SVR generally converts training data from the original dimension to a higher dimension to effectively find the optimum hyperplane [22].

  ⇒ The accuracy of air temperature was improved from the Local Data Assimilation and Prediction System (LDAPS) model in Seoul, South Korea, by using SVR and a linear regression model, finding that SVR showed higher correction accuracy than the linear regression model.

- Multi-model ensemble (MME): MME is used to combine multiple machine learning models. Generally, this approach enables users to achieve higher accuracy and robustness when compared to using a single model and serve to improve NWP model-derived daily maximum and minimum air temperature [22].

Challenges

- There are currently several different machine learning algorithms being used in temperature bias correction. However, improving models accuracy remains a challenge, especially when the forecast time horizon increases.

- An individual machine learning algorithm cannot reduce the NWP model bias consistently and effectively due to the complex connection between the atmosphere and the ground [25].

- In contrast to global models that currently possess a high capacity to forecast precipitation based on synoptic scale features, high-resolution regional models have the advantage of capturing locally heavy rainfall. Nevertheless, for cases with multiple precipitation from tropical cyclones (TCs), it is a challenge for all models, whether they are global or regional, to provide reliable guidance on TC rainfall [26].

- There may be a trade-off between sharpness and reliability when evaluating forecasts after bias correction. Additionally, the evaluation of forecast reliability shows that while bias correction improves the forecast reliability, more improvement is still required [27].

Impact

Generally, effect analysis of relevant meteorological (such as mean, maximum, and minimum temperatures, rainfall, and relative humidity) and geographical (such as latitude, longitude, and elevation) variables improves air temperature forecasting accuracy. In future research studies, the use of feature selection techniques, like recursive feature elimination, and correlation coefficients, will aid in selecting the optimal input variables for air temperature and precipitation prediction.

6.2.3 Convolutional Neural Networks for Wind Forecasting

High-resolution visual data can now be used for weather monitoring and forecasting thanks to development of machine learning. Accurate wind forecasting is particularly important for a variety of economic, business, and management sectors. One of the most well-known natural disasters is typhoon. Typhoon can cause huge damage to agriculture, buildings, etc. and further affect the economy of human society. How to accurately predict wind (wind speed, wind direction, type of wind, etc.) has become a hot and worthy research topic. In this trend, we will introduce different methods of using Convolutional Neural Networks (CNN) to predict wind.

Facts

- Typhoon is one of the most frequent disasters. It will not only cause powerful damage, but also cause casualties. For example just affected by Typhoon Haiyan, 6,352 people were confirmed dead, 1,771 people were missing, causing economic losses of 2.98 billion Dollar. [28]

- Traditional numerical forecast models based on fluid mechanics have difficulty in predicting the intensity of typhoons. Forecasts based on statistics and machine learning fail to take into account the spatial and temporal relationships among typhoon formation variables leading to weaknesses in the predictive power of this model. [29]

- Typhoon messages are a type of observational data that record the occurrence, development, end status, intensity and path prediction of typhoon. According to the data in the typhoon messages, the typhoon occurrence process can be analyzed and the quality of the typhoon initial value in the future climate model simulation can be improved The problems in the reception of global typhoon messages are: 1) Typhoon pre-warning messages High latency caused by passive reception; 2) The problem of incomplete typhoon message data. [30] If the complete set of typhoon message data cannot be collected, this data will not be used for scientific research.

- Global mean land surface temperature and sea surface temperature (SST) have risen significantly over the past half century, and the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) concluded that most of the global surface temperature increase was very likely due to the observed increase in anthropogenic greenhouse gas concentrations. Given the great societal and scientific concerns, the impact of global warming on tropical cyclone (TC) activity has been the subject of considerable investigation in recent years. While progress has been made in assessing the influence on TC intensity and rainfall, relatively little is known about the possible change of TC tracks in a warming climate. [31]

Key Drivers

- In order to deal with the complex typhoon climate problem, multiple models with different advantages can be used to construct a composite model. Such as the three components of the hybrid model, the 3D-CNN (is used to analyse atmospheric variables in 3-dimensional space); 2D-CNN(is used to analyse those at the sea surface); and, LSTM (is used to capture the temporal correlations) are combined to collate data which is used to analyse the relationship between spatio-temporal phenomena and the variables of typhoon formation and intensity. [29]

- On the basis of the original message collection process, an additional typhoon recognition module implemented through deep learning is added to actively acquire and receive typhoon messages to complete the purpose of collecting and supplementing the collection of messages in advance. [30] That is, through the neural network to improve the timeliness of the collected messages and increase the types of collected messages.

- The current typhoon monitoring and prediction are mainly based on simulation with meteorological data; the accuracy still needs to be improved. Nowadays, the technology of Internet of Things (IoT) and remote sensing technology become more and more closely linked; many IoT systems in smart cities' can obtain high-resolution remote sensing image data. Therefore, it is possible to use urban IoT system to realize the early warning of typhoon. [32]

Challenges

- Based on the existing methods, it is still difficult for us to predict the long-term future for wind speed and maintain a high accuracy rate. [33]

- The function of the sensor is limited, people cannot get the ideal data for analysis

- Typhoons can be roughly divided into two categories: One is the typhoon with an eye, and the other is typhoon without eyes. Eye regions of typhoons are more obvious on meteorological satellite images. In many cases, images of non-eye typhoons have no obvious features. [32]

- People have not explored some challenging topics such as anomaly detection, partial annotation detection and transfer learning (e.g. to satellite imagery). [34]

Impact

The reliability and practicality of CNN have been fully proved in the analysis and prediction of one of the most frequent and destructive types of extreme weather - Typhoon. Due to the possible combination of CNN with other DNN structures, the application of CNN can be widely extended. Therefore, CNN plays a vital role in the prevention of huge disasters such as floods and tsunamis caused by typhoons. Additionally, the use of neural networks to forecast typhoons will have more realistic economic and social benefits as the concepts of smart cities and IoT become more mature and popular.

6.2.4 Prediction of pluvial Flooding Events

In recent years, the number of floods caused by intense rainfalls has increased steadily. This is mainly due to climate change and urbanisation. We became aware of the fact that this is an existential threat to us again this year in Germany. More than 160 people died in Rhineland-Palatinate and North Rhine-Westphalia alone. There is also criticism of the incompletely functioning early warning system for extreme climate events. The predictions of climate researchers are that such events may occur more strongly and more frequently in the future due to global warming. Warmer temperatures cause more water to evaporate and the air to absorb more moisture. A decreasing jet stream ensures that weather conditions stay longer in one place[35]. Therefore, more rain can fall in these areas. To predict such events, a distinction is made between long-term prediction and short-term prediction. This is done because there is no predictor that can make a good prediction at all times. Long-term forecasting is needed to take preventive measures. Here, risk areas are to be identified and as few as possible inhabited or protected. The short-term forecast is used to warn the inhabitants of an impending flood and to bring them to safety.

Facts

- Flooding due to heavy rainfall is one of the most common types and has high priority in the development of early warning systems[37].
- About 40 % of the damages and associated economic losses in the UK are from flooding[38]. In China 98% of the cities are exposed or vulnerable to frequent floods[39].
- Initial case studies have shown that Ml models can have a 5 % - 15 % higher accuracy and true positive rate and can thus improve existing warning systems[40].

Figure 6.2: Flooding in Meißen, Germany [36]

- Ml algorithms are usually more efficient and easier to use than traditional computationally intensive hydraulic models[40].

Key Drivers

- Data science and machine learning are constantly being developed and, thanks to powerful hardware and effective algorithms, are very well suited to complex problems such as this one.
- Increasing amounts of data are being generated, such as data from satellites, which can lead to better performance[41].
- As the problem of flooding continues to grow, policy makers continue to promote and invest in it.
- As more and more people are moving to the city and these conurbations are affected much more frequently due to dense development. In addition, riverine development often contributes to making the situation worse.

Challenges

- In regions that have not been so badly affected by floods, there is not much data.

- It is not yet possible to foresee exactly what influence climate change will have in the future. This makes it particularly difficult to make long-term predictions.

- In most cases, no universal solution can be found, but one must adapt to individual areas.

- Scientists from several disciplines have to work together, which can be a challenge and can affect the outcome.[42].

Impact

Machine learning can be used to build more accurate systems, as demonstrated by an example in Shenzhen [40]. A promising point is the further development of hybrid models, with which better results were achieved in both short-term prediction and long-term prediction [42]. In addition, it makes sense to use radar and satellite data, through which an increase in performance can be achieved [41]. With a lot of data from different sources, the positive influence of decomposition methods is noticeable[42]. To develop a well-functioning system, scientists from many fields need to collaborate and share their expertise. New knowledge from the fields involved can make these predictions more accurate.

6.2.5 Segmentation of Atmospheric Rivers and Tropical Cyclones using Deep Learning

Tropical cyclones (TC), extra-tropical cyclones (ETC) and atmospheric rivers (AR) are severe and powerful weather events. Depending on their size and shape they may lead to extreme rainfalls, hurricanes or typhoons. Traditional methods to detect these phenomena rely on sequentially processing of the same data to identify each event category (TC, ETC or AR). However, it would be much more effective to detect all types of extreme weather events based on the characteristics and trends that exist in multi-variate climatic datasets. Deep learning can be used as a sensor and an automated extreme weather tracker that relies on spatiotemporal models, instead of thresholds, in climate model simulations. With the help of deep learning, scientists can better investigate the environmental factors that control the frequency, intensity and location of these extreme weather phenomena and how they would vary in different climate change scenarios due to global warming. [43] This section is assigned to Sefa Ekin Yilmaz.

Facts

- Traditional extreme weather detection methods rely on subjective, arguably arbitrary thresholds that may change under different climate change scenarois due to global warming [43].

- One obstacle of choosing different thresholds or variables in traditional detection methods would be that there are large discrepancies between their outputs even for the same type of pattern or event [44].

- Deep learning models can be applied to different data sets without tuning, since they do not rely on threshold conditions unlike the traditional detection methods, using these models would be beneficial under climate change conditions [44].

Key Drivers

- Recent research has shown that deep learning can be applied to identify the type (classification), spatial extent (location) and pixel-level (segmentation) masks of weather and climate models [21].

- Yet, key patterns of extreme-causing weather events are not very well understood, making them difficult to be forecasted beforehand. Since machine learning approaches have yielded promising results in representing physical processes in the ocean and atmosphere, it is believed that these techniques have the potential to make important improvements in climate modelling and long-term climate projections.[45]

- Deep neural networks are capable of learning high-level representations of patterns from labeled data. Bearing in mind that many factors affect the occurancy of these extreme weather events, using such techniques would produce reliable results for extracting the patterns that lead to such events.[21]

Challenges

- One of the most important requirements for the success of supervised deep learning models is the quality and reliability of the data labelled by experts. There is a lack of crucial experts-labeled datasets in fields of weather and climate science. [44]

- In extreme weather events segmentation problems the data set is usually imbalanced, since in the images there are a lot more background pixels than the pixels of AR, ETC or TC [43].

- There are also many physical factors that have impact on the occurancy of tropical cyclones, extra tropical cyclones or atmospheric rivers and these factors vary for each event. For example atmospheric rivers are correlated with the vertical integration of water vapor, while tropical cyclones correlates with sea level pressure, near surface wind and upper troposphere temperature, which make identification of these events even more challenging. [46]

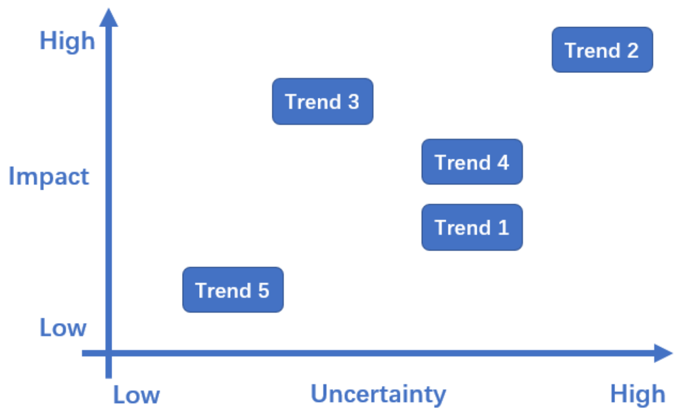

Figure 6.3: Comparison trends in trend matrix

Impact

Keeping in mind that there are large discrepancies between different traditional extreme weather event prediction methods and additionally since these methods rely on different threshold conditions, the use of deep learning methods to segment these events is beneficial especially to be able to cope with them in different climate change scenarios due to global warming.

6.3 Conclusion

In the previous chapter five trends was introduced. Five trends are compared through two dimensions: uncertainty and impact. Uncertainty represents the problems and challenges encountered by the trend. And impact is his effects or benefits on the future. The result is showed in figure 6.3.

In recent decades, the monitoring of the earth, environment and weather systems has made rapid progress. Meteorologists are increasingly turning to artificial intelligence to improve weather modeling, from predicting short- and long-term patterns to predicting life-threatening extreme weather events. A total of five trends have been proposed in this article, involving the prediction of extreme hail weather, precipitation prediction, wind forecasting, flood event prediction and segmentation of weather events.

Every trend we propose plays an important role in the field of extreme weather forecasting, and they also face various challenges. There are some common challenges, which are even the challenges encountered by all artificial intelligence models. For example: the accuracy of the collection of a large amount of data, the accuracy of long-term forecasts and the joint cooperation between interdisciplinary.

Since this industry is still in its early stages, there are many challenges in the future. The core of using artificial intelligence to predict extreme weather is to establish a reasonable and effective mathematical and data model, and to ensure the stability of the algorithm.

References

[1] Donald J Wuebbles et al. "Climate science special report: Fourth national climate assessment (NCA4), Volume I". In: (2017).

[2] Adam B Smith. "2020 US Billion-Dollar Weather and Climate Disasters—In Historical Context". In: *101st American Meteorological Society Annual Meeting*. AMS. 2021.

[3] F Kreienkamp et al. "Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021". In: (2021).

[4] Richard B Alley, Kerry A Emanuel, and Fuqing Zhang. "Advances in weather prediction". In: *Science* 363.6425 (2019), pp. 342–344.

[5] David Rolnick et al. *Tackling Climate Change with Machine Learning*. `https://arxiv.org/pdf/1906.05433.pdf`. [Online; accessed: 27-August-2021]. 2019. arXiv: `1906.05433 [cs.CY]`.

[6] Ryan Jewell and Julian Brimelow. "Evaluation of Alberta Hail Growth Model Using Severe Hail Proximity Soundings from the United States". In: *Weather and Forecasting* 24.6 (2009), pp. 1592–1609. DOI: `10.1175/2009waf2222230.1`.

[7] A. Svaldi. "Damage from last year's massive front range hail storm cost 2.3 billion - 900 million more than first estimated". In: (2018). URL: `https://www.denverpost.com/%202018/05/07/2017-front-range-hail-storm-damage-cost/`.

[8] Tomáš Púčik et al. "Proximity Soundings of Severe and Nonsevere Thunderstorms in Central Europe". In: *Monthly Weather Review* 143 (Sept. 2015), p. 150929114615006. DOI: `10.1175/MWR-D-15-0104.1`.

[9] David Šaur. "Evaluation of the Accuracy of Numerical Weather Prediction Models". In: *Advances in Intelligent Systems and Computing* 347 (Jan. 2015), pp. 181–191. DOI: `10.1007/978-3-319-18476-0_19`.

[10] MESH Raw and MESH Filtered. "76 Exploration of the NSSL Maximum Expected Size of Hail (MESH) Product for Verifying Experimental Hail Forecasts in the 2014 Spring Forecasting Experiment". In: ().

[11] David Gagne et al. "Storm-Based Probabilistic Hail Forecasting with Machine Learning Applied to Convection-Allowing Ensembles". In: *Weather and Forecasting* 32 (Aug. 2017). DOI: `10.1175/WAF-D-17-0010.1`.

[12] Alexandru Niculescu-Mizil and Rich Caruana. "Predicting good probabilities with supervised learning". In: Jan. 2005, pp. 625–632. DOI: `10.1145/1102351.1102430`.

[13] Amanda Burke et al. "Calibration of Machine Learning-Based Probabilistic Hail Predictions for Operational Forecasting". In: *Weather and Forecasting* 35 (Nov. 2019). DOI: `10.1175/WAF-D-19-0105.1`.

[14] David John Gagne II et al. "Interpretable deep learning for spatial analysis of severe hailstorms". In: *Monthly Weather Review* 147.8 (2019), pp. 2827–2845.

[15] Farha Pulukool, Longzhuang Li, and Chuntao Liu. "Using Deep Learning and Machine Learning Methods to Diagnose Hailstorms in Large-Scale Thermodynamic Environments". In: *Sustainability* 12.24 (2020), p. 10499.

[16] Iksha Gurung et al. "Deep feature extraction and its application for hailstorm detection in a large collection of radar images". In: *Signal, Image and Video Processing* 13 (Apr. 2019). DOI: `10.1007/s11760-018-1380-z`.

[17] Melinda Pullman et al. "Applying Deep Learning to Hail Detection: A Case Study". In: *IEEE Transactions on Geoscience and Remote Sensing* PP (Aug. 2019), pp. 1–8. DOI: `10.1109/TGRS.2019.2931944`.

[18] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: June 2016, pp. 770–778. DOI: `10.1109/CVPR.2016.90`.

[19] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: `1409.1556 [cs.CV]`.

[20] Fabian Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (Jan. 2012).

[21] Yunjie Liu et al. *Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets*. 2016. arXiv: `1605.01156 [cs.CV]`.

[22] Dongjin Cho et al. "Comparative assessment of various machine learning-based bias correction methods for numerical weather prediction model forecasts of extreme air temperatures in urban areas". In: *Earth and Space Science* 7.4 (2020), e2019EA000740.

[23] Donghyuk Kum et al. "Projecting future climate change scenarios using three bias-correction methods". In: *Advances in Meteorology* 2014 (2014).

[24] B. Cui et al. "Bias Correction for Global Ensemble Forecast, Weather and Forecasting". In: 27 (2012), pp. 396–410.

[25] Jui-Sheng Chou and Anh-Duc Pham. "Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength". In: *Construction and Building Materials* 49 (2013), pp. 554–563.

[26] WC Woo et al. "Challenges and advances related to TC rainfall forecast". In: *Third Int. Workshop on Tropical Cyclone Landfall Processes*. 2014.

[27] Louise Crochemore, Maria-Helena Ramos, and Florian Pappenberger. "Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts". In: *Hydrology and Earth System Sciences* 20.9 (2016), pp. 3601–3618.

[28] James Daniell et al. "CEDIM Forensic Disaster Analysis". In: (2013).

[29] Rui Chen et al. "A hybrid CNN-LSTM model for typhoon formation forecasting". In: *Geoinformatica* 23.3 (2019), pp. 375–396.

[30] HAN Rui et al. "Global Typhoon Message Collection Method Based on CNN-typhoon Model". In: *Computer Science* 47.11A (2021), pp. 11–17.

[31] Ruifang Wang, Liguang Wu, and Chao Wang. "Typhoon track changes associated with global warming". In: *Journal of Climate* 24.14 (2011), pp. 3748–3752.

[32] Eric Ke Wang et al. "Intelligent monitor for typhoon in IoT system of smart city". In: *The Journal of Supercomputing* 77.3 (2021), pp. 3024–3043.

[33] Gwo-Fong Lin and Lu-Hsien Chen. "Application of an artificial neural network to typhoon rainfall forecasting". In: *Hydrological Processes: An International Journal* 19.9 (2005), pp. 1825–1837.

[34] Evan Racah et al. "ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events". In: *arXiv preprint arXiv:1612.02095* (2016).

[35] Michael E. Mann et al. "Projected changes in persistent extreme summer weather events: The role of quasi-resonant amplification". In: *ScienceAdvances* 4 (2018).

[36] Pixabay License. *Line Spacing in LaTeX documents*. [Online]. `https://pixabay.com/de/images/search/hochwasser-elbe-mei/`.

[37] R.H. Falconer et al. "Pluvial flooding: new approaches in flood warning, mapping and risk management". In: *Journal of Flood Risk Management* 2 (2009), pp. 198–208.

[38] I. Douglas et al. "Urban pluvial flooding: a qualitative case study of cause, effect and nonstructural mitigation". In: *Journal of Flood Risk Management* 3 (2010), pp. 112–125.

[39] Yong Jiang, Chris Zevenbergen, and Dafang Fu. "Can "Sponge Cities" Mitigate China's Increased Occurrences of Urban Flooding?" In: *Aquademia* 1 (2017).

[40] Qian Ke et al. "Urban pluvial flooding prediction by machine learning approaches –a case study of Shenzhen city, China". In: *Advances in Water Resources* 145 (2020).

[41] M. Grecu and W. Krajewski. "A large-sample investigation of statistical procedures for radar-based short-term quantitative precipitation forecasting". In: *J. Hydrol.* 239 (2000), pp. 69–84.

[42] Amir Mosavi, Pinar Ozturk, and Kwok-wing Chau. "Flood Prediction Using Machine Learning Models:Literature Review". In: *Water* 10 (2018).

[43] Mayur Mudigonda et al. "Segmenting and tracking extreme climate events using neural networks". In: *Deep Learning for Physical Sciences (DLPS) Workshop, held with NIPS Conference*. 2017.

[44] Karthik Kashinath et al. "ClimateNet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather". In: *Geoscientific Model Development* 14.1 (2021), pp. 107–124.

[45] Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. "Analog forecasting of extreme-causing weather patterns using deep learning". In: *Journal of Advances in Modeling Earth Systems* 12.2 (2020), e2019MS001958.

[46] Adam Rupe et al. "Towards unsupervised segmentation of extreme weather events". In: *arXiv preprint arXiv:1909.07520* (2019).

Chapter 7

Collective Decision Making for the Environment

Deuschl, Xaver
Forster, Kerstin
Gomez Ryfka, Adrian
Karanfilovska, Marina
Ni, Kaichen
Nüsslein, Stephan
Shobayo-Eniola, Abidemi
Zhu, Yuqicheng

Abstract

Different ways in which machine learning can be used to avert the climate crisis have been proposed. One of the most critical aspects of climate change is making the right decisions that lead to more climate-friendly actions, prevention and mitigation. The challenge of such a global problem, where the consequences are not so direct and obvious, is to find agreements between the different actors and interests involved. This trend report is based on a review of state of the art available documentation and also reflects own research and assessment. The focus lies on the use of machine learning techniques to improve climate-related decision making. Our investigations concur in two key requirements. First, objective and accurate models to act with precision, minimize costs and consequently also have more persuasion power, essentially via Integrated Assessment Models. Second, the will to apply this knowledge to individual decisions by nudging and globally through agreements. Hence, appropriate education and negotiation methods are necessary, such as Serious Games and Agent Based Models.

7.1 Introduction

Technology by itself will not stop climate change. Technological advances must target critical aspects of climate change accurately, allowing us to reduce CO_2 emissions from an individual level up to larger scaled organizations such as companies and countries. To promote large-scale adaptation of climate-friendly technologies and CO_2 reduction measures, adequate policies and, consequently, widely spread environmental awareness among society are essential. Can data science help achieve these goals?

This is where data-driven decision making comes into play. Data-driven decision making refers to the practice of making decisions based on the analysis of data rather than on intuition or observation alone. In the context of climate change, data-driven decision making can be highly useful for both individual and collective decisions [1].

Each one of us is constantly confronted with decisions affecting emissions, but may lack the knowledge of which decisions have the greatest impact on reducing it. Basing our decisions on data, e.g. in the form of nudging, can help maximize their positive effect on the climate [2].

Collective decisions made, for instance, by governments, businesses, NGOs, or intergovernmental organizations comprise many kinds of action and often involve multiple stakeholders with different priorities and motivations. To successfully analyze and process all the variables involved in the context of climate change and make decisions that yield mutually beneficial outcomes, mathematical methods from fields such as policy analysis or game theory as well as accurate climate models supported by machine learning are crucial to guide the decision-making process [2].

In a more profound way, both individual and collective decisions are also shaped by climate education, which helps people across societies understand and address the repercussions of climate change. In this manner, education raises awareness of climate issues and how to tackle them – and thereby positively influences climate-related decision making of both individuals and organizations. Educational methods incorporating data-driven insights have proven particularly effective due to their personalized and interactive nature [2].

This report presents four emerging trends in the field of data-driven decision making. It is structured as follows: in Chapter 2, the results of our research are demonstrated. To conclude, all trends are assessed and compared in terms of impact and uncertainty in Chapter 3 and an outlook on future developments is provided.

7.2 Trends

In the following section different trends are presented. First human decision making and the trend for more personalization using AI is discussed. This might be used to reduce the CO_2 emissions by Nudging. Next AI improvements for Climate Models are discussed. The third trend is the use of AI for policy assessment. Climate models and policy assessment are important for designing good policies that reduce CO_2 emissions. Finally Serious Games are presented. These might help to educate about climate change.

7.2.1 Nudging

Many CO_2 emissions are related to consumption habits. By changing these, a reduction of emissions can be achieved [3].

To describe human behaviors we have to use some type of model. Economic models characterize the actor as rational, egoistic and with constant preferences. But it is clear from observations that humans do not behave this way. A better way to describe human decision making is a model with two Systems, System 1 and System 2.

- System 1 works automatic and fast, and does not require deliberate control.
- System 2 is used for more mentally taxing tasks. This system is slow and requires concentration.

Usually we associate human decision making with conscious and logical decisions and deliberate control of actions. This is associated with System 2. But actually System 1 does most of the decisions and is in control of the impressions and feelings. This makes clear that humans do not always make rational decisions[4]. This makes it possible to influence these decisions based on these observations. One way to influence human decisions is nudging. This describes changes to the choice architecture without mandates, bans or financial incentives. The main idea is to make some option easier, but still allow for different choices to preserve liberty. Basic techniques are the use of default rules or simplification in forms, which make it easier to do a certain thing. Other techniques are the use of social norms, where it is emphasized what most people do and informing people about the consequences of their choices. To ensure that these work as intended, empirical tests are needed[5].

Facts

- Advertisement often operates based on emotions[6]. In user interfaces Dark Patterns are used to exploit human behaviour [7].
- Some type of choice architecture is always present, as human decisions are always influenced by the surroundings. So new nudges usually replace preexisting nudges[5].
- Nudging can also be done in a digital environment. The idea is to personalise it and make it context aware to increase the effectiveness. This requires processing of various different information sources to create the user profile, process the context, select the relevant information and finally design and present the nudge [8]. The process is illustrated in figure 7.1.

Key Drivers

- Recommendation systems are used by companies for their e-commerce and many other purposes [9]. Also there is a growing trend toward personalization [10].
- Personal digital assistants like Siri, Alexa or Google Now which are becoming more and more widespread [11]. This will make recommendaton systems more important for everyday decision making.
- One classic example for the usage of AI for a recommendation system is Netflix. Here they use different algorithms to compile personalized recommendations. These are also further improved using A/B-testing, where some randomly chosen users get the recommendation from an altered algorithm [12].
- To make personal decisions more ecological the reduction of CO_2 emissions has to be an additional objective. Business sustainability and COVID mitigation requirements where included into a recommendation system alongside customer utility [13]. A different proposal is to use AI to nudge people to more healthy food by recommending healthier recipes in recipe websites, by using more appealing images for healthier options [14].
- Feedback can be used to help customers reduce their energy consumption [15].
- IOT technology can add situational awareness, this means knowledge of the current environment like weather, or current traffic. By personalization the information can be relevant for the person, so take personal preferences into account[16].
- People are more likely to use the bike if they have an appropriate and safe infrastructure. Apart from the reduced CO_2 emissions and other ecological benefits active transport (walking and cycling) has health benefits as well. There is already work done on how to facilitate AI and Data science to improve cycling safety and usage[17, 18].

Challenges

- Preservation of privacy limits that certain data are only processed on the user's device, or it has to be ensured that the transmitted data does not violate the privacy of the user. This is especially difficult for complex systems as AI, because user data that is used to improve the system can cause a privacy infliction[8].
- There has been some debate about the ethics of nudging. From these discussions some aspects should be considered for the implementation of nudges [19]. A further aspect is that humans are not rational actors. As such, humans might need help with their decision making [4].

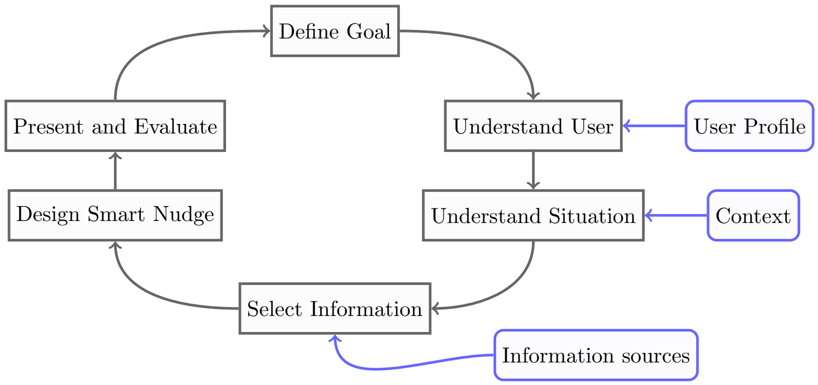

Figure 7.1: Process of designing a Smart Nudge (Simplified from [8])

- To be impactfull the nudging needs to be widespread and receive support by influential actors, like the popular search engines.

Impact

The effect of nudging is only minor compared to other interventions. Also the effect of nudging is only limited and sometimes leads to less ecological behavior in other aspects. Nudging can lead to diminishing support for other measures, as long as people are not told about the limited potential of nudges [20]. It is easy to conceive that the notion of AI as a solution to climate change has the same effect of reducing the support for other more effective means of CO_2 reduction. It is probably equally important in this case to communicate the limitations of CO_2 emission reduction achievable with AI. By this we can avoid possible additional emissions due to a lack of employment of more effective means to combat climate change.

7.2.2 Improving climate models

Extensive and accurate models that describe the physical processes of climate change and its impacts and causes, are essential to understand the implied risks and how to avoid or mitigate them. The same way that it is not possible to win a chess tournament without knowing its rules, we need the background information about climate change to be as precise as possible when making decisions or designing policies.

Facts

- Integrated Assessment Models (IAMs) are being used to model physical, economic and social aspects of climate change and their interaction with each other. Some of them play a prominent role in the IPCC assessments and are official references for supranational organizations such as the United Nations [21]. Machine Learning's accuracy and ability to deal with huge amount of data, makes it a very powerful and adequate tool for this purpose [2].

- Because of the complexity of the dealing within the various biogeochemical and socioeconomic components, a number of quantitative models have been developed to study those relationships and the effects of the public policies regarding of future climate change [21].

- Responding to the climate change requires rapid and effective decision-making by groups at multiple levels – communities, unions, NGOs, businesses, governments, intergovernmental organizations, and many more. These decisions are related to multiple stakeholders with different goals and priorities, necessitating difficult trade-offs [2]. Therefore, we need unbiased and optimal criteria to minimize the social and economic costs, thus being able to convince as many people as possible.

Key Drivers

- Scientists are the most obvious actors in this topic. From biologists, physicists, engineers and computer scientists to economists and even psychologists. They are the experts involved in the task of modeling climate and people's behavior. There might be groups willing to contribute out of altruism and academic purposes, but such an expertise and hard work also deserves economic retributions. This is where the next groups come into play.

- Companies tend to adapt to satisfy their customer's needs. However, most of us are not aware of our carbon footprint when we make decisions such as which phone we buy for example. While this is the case, companies will not have strong incentives to behave sustainably. This is one of the main reasons why individual decision making is so important and can be supported with the previously discussed methods.

- Nevertheless, most of the decisions regarding climate change, concretely regulations and development plans, are of governmental responsability. In our opinion, they should take the lead and provide the experts with economical finance, as well as seriously taking the experts results into consideration when developing related regulations and plans in a symbiotic manner.

Challenges

- The potential for catastrophic climate change impacts is among the most important concerns about climate change. Designing economic cost of such catastrophical events exists in IAM models, but there is a lack of basis in order to determine the size, timing, or probability of the tipping points of such events or the damage caused on the economy. In this case, it is necessary to collect the expert opinions for the characterization of the functional form and uncertainty of these tipping points [21].

- Moreover, obtaining the relevant data is not trivial. Many non democratic countries may not have incentives to share information about their habits, for example.
- Lack of scientific knowledge of complex physical processes with many interdependant variables, such as some natural processes.
- The main focus of the to date IAMs is the efficiency. Lack of equality consideration within and between nations in these models can be problematic, as those issues are often topics of political debate for climate change. Treatment of socioeconomic aspects at regional, national and international level should be considered in IAMs [21].
- Additionally to fair treatment of the countries, intertemporal and intergenerational equality should be adressed so the appropriate discount rate of implementation of policies for mitigation can be designed [21].
- IAMs to date simplify assumptions about risk attitudes and do not cover the alternative assumptioms at all, which could lead to understating the society's degree of risk aversion to poor outcomes and resulting mitigation could be less than desired [21].

Impact

Extensive and accurate information is key to tackle climate change. Understanding its exact causes and consequences is the only way to effectively act in consequence. Another benefit is that we create realistic, practically non-skewed information to help us find optimal solutions in policy assessment. The challenges described show some uncertainty in the impact, specially since having the right information and models does not guarantee using them wisely. When dealing with a global problem, it is important to negociate with objective, unbiased data in order to embrace all political views and cultures. This is the most promising negotiation method to convince as many people as possible and collaborate for the best results, which leads us to the next section.

7.2.3 Policy Assessment

Assessing policies is a part of policy advisory. There is an increasing need for advisory due to the fact that the problems that have to be solved in politics, especially regarding the climate change, are increasingly more complex or even unsolvable. These issues are often referred to as "Wicked Problems" [23].

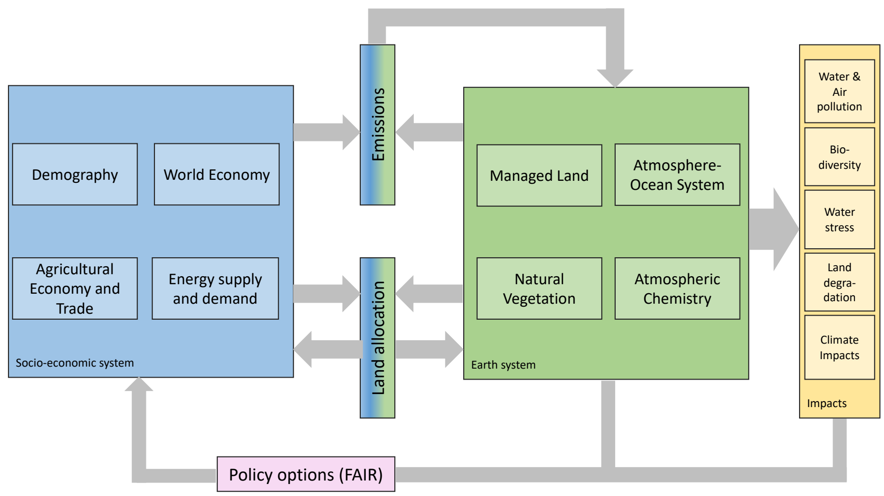

Figure 7.2: Diagramatic Examples of RCP Integrated Assessment Models [22]

Policy advisory is more a continuing support of the decision process and not only the handover of a ready-made solution. Therefore, policy advisory itself has to be understood as a process [24]. Good policy advisory should be fact-supported and therefore not based on ideological prejudices or theoretical debates, but grounded on practical observations and empirically evaluations [23]. The advice given by scientists should be both epistemologically and politically robust. "Evidence-based policy-making" is the idea that political decision-making can find more effective solutions to the problems it faces if it uses scientific findings. This rational-technocratic approach also has its limitation, namely that in this process not only knowledge has to be managed but also the limits of knowledge and uncertainty [23]. Also, real world limitations, like the limited knowledge or time, make it impossible to reach ideal rationality [24].

The results of policy advisory are usually not binding. But sometimes they are integrated into a formal context with some bindingness [23]. Juristically, there is also the problem that the scientific results, that are necessarily not democratic, have to be embedded into a political process that draws it legitimacy from the democratic process. On the other hand there is also an argument to be had that there is also an obligation to include expertise into the decision process [25]. Nevertheless, with the development of techniques like natural language processing (NLP) or agent-based models there is huge potential in analyzing previous policies and the effects of new ones using machine learning.

Facts

- In the fight against global warming, politics must also play its role and make the right changes to the law.
- The areas where our environment is polluted by human intervention and where we need to take action are mostly known. In addition, it is possible to define different tipping points that should not be exceeded. Where these are located is not yet fully explored but there are already different approaches to define them, as described in [26].
- Various legislative tools, that can be market-based, law-based or regulatory are available. Examples could be subsidies, taxes, green certificates or tradable permits described in [26] extensively.
- More and more data on the behaviour of different actors such as individuals, communities, associations, unions, NGOs, governments and many more are widely accessible and continue to be collected (e.g. essential for developing behavior rules of the individual agents).

Key Drivers

- (Multi) Agent-based models (ABM's) in combination with Reinforcement Learning like in [27] or [28].
- Combinatorial optimization tools [29].
- Integrated assessment models (IAMs) presented in [30] and [21].
- Data-driven modelling frameworks for learning behavioural rules and necessary attributes that need to be initially assigned to the agent-based model presented in [31] and [32].

Challenges

- So far, only a relatively small number of agents can be modelled. This means that the results are not yet transferable to nations or unions on a large scale.
- The design of attributes, learning behaviour and rules for the agents and the environment in which they operate is usually still designed by experts, which is why this is very expensive and time-consuming [2].
- In some urgent cases (like the pandemic), machine learning models could be unreliable due to the lack of historical data. Who should take responsibility for the decisions made, which are becoming increasingly difficult due to the rapidly progressing climate change?
- The simulations of different policies can only be as good as the different climate models that are included. Therefore, it is important that as in chapter 7.2.2 the reliability of climate models increases.

- Many different stakeholders and interests play a role in the development of legislation. The question arises to what extent the AI solution is included in the decision-making process, even if it is the optimal or most rational solution.

Impact

The trend described can help to analyze policies for global warming more systematically and thus to make better political decisions for the future. Just understanding how different approaches and versions of a policy can play out can be enormously helpful. This trend is of great interest as the decisions to be made are influenced by more and more constraints and interferences, making it difficult for legislating governments or associations to keep track of them. A good example of this and the possible impact is described in the paper [27], where it was shown that with carbon prices in various amounts, land consumption and CO_2 emissions could be greatly reduced despite the rising net income of the simulated farmers. However, there are also limits to the described approach. ML is merely a tool that provides a data-based and hopefully near-optimal basis for decision-making (Pareto-optimal). How this basis is dealt with, whether it is observed or not, remains up to the politicians, which can greatly reduce the influence of the trend described and is a factor of uncertainty. No matter how rational a decision seems to be, this unfortunately does not mean that it will always be taken.

7.2.4 Serious Games for Climate Education

Serious games, or games used for purposes other than mere entertainment, are nowadays applied to a broad spectrum of areas, including military, government, educational, corporate, or healthcare [33]. As opposed to traditional learning methods, serious games enable their players to learn through first-hand experiences – and thereby target affective outcomes like motivation, attitudes, and values. They allow for visioning, experiencing the consequences of actions and taking on various roles and perspectives, which makes them particularly apt for climate change education [34].
A major difficulty in communicating climate change is the high complexity of the global atmospheric systems involved. Climate games manage to bring this complexity to a hands-on level by providing a safe environment to explore different actions and decisions. They are used to teach climate change mitigation, adaptation, or climate science in general and can be adjusted to a variety of target audiences, ranging from children to policymakers [35]. Serious games present various opportunities for data science and machine learning (ML), including the creation of engaging visualizations [36], realistic non-player characters (NPCs) [37], inference engines generating new knowledge, or game decision support systems recommending user actions considering the current status of the game [33].

Facts

- Serious games have experienced a continuous rise in popularity: since 2011, over 62,000 publications containing the key word “serious games” have been made made until 2018, numbers increasing each year [35].

- Global market revenue based on serious games is expected to grow from USD 3.5 billion in 2018 to 24 billion in 2024 [38].

- The first serious games relating to climate change date back to the 1980s and began as board games modeling increases in atmospheric CO_2 levels [34]. Since then, climate change games have evolved greatly in technical sophistication: today’s trends include blending digital and real-world elements and incorporating social networking [34].

- Examples for state-of-the-art climate games are ”Mission 1.5” developed by the United Nations Development Programme (UNDP) [39], ”Climate Kids” developed by NASA for upper-elementary-aged children [40], or ”SIM4NEXUS”, a Horizon2020 project targeted specifically towards individuals in decision-making and decision-influencing positions [33].

Key Drivers

- The number of people playing video games has been on a constant rise and reached 2.69 billion gamers globally in 2020, which makes up over one third of the world population [41]. As a result, gamers represent a large potential audience for raising awareness and promoting engagement with regards to climate change.

- Progress in visualization and interaction hardware devices and software systems enables the development of increasingly advanced serious games using, for instance, augmented and virtual reality (AR/VR) techniques or artificial intelligence (AI) [37].

- The German government has proposed a digitalization strategy aiming to provide Germany with high-speed net until 2025 and to bring AI research and application in Germany to a globally leading level [42]. Both strategies encourage the development and usage of more numerous and technically sophisticated serious games.

Challenges

- Balancing complexity and simplification of climate-related phenomena, as well as the degrees of reality and fun elements, is crucial for the success and impact of climate games. However, achieving the right balance is often challenging [43].

- While evidence exists for the positive short-term impact of climate games on the players' attitudes towards climate change [44], research on long-term behavioral changes remain inconclusive. Conducting comprehensive studies in this regard therefore remains a challenge for future research [34].
- Using realistic visualizations in serious games to influence people's values and opinions raises ethical dilemmas, which must be considered carefully prior to commercialization of a climate game [45].

Impact

Despite posing serious challenges, serious games present a great opportunity to raise public awareness, understanding and engagement regarding climate change. With the key ability to foster motivation and empathy towards climate issues, serious games are crucial to shaping a new culture of responsibility and action in the fight against global warming. Technological advances enable the development of yet more engaging and diverse games, and thereby the expansion to a larger number of target audiences and sustainability topics. ML in particular can be applied in a variety of ways to increase the effectiveness of serious games and, ultimately, their impact on climate-related decision-making.

7.3 Conclusion

This report highlights the importance of addressing the climate change impacts in the decision-making process for different target groups of people or institutions. We presented different trends in the context of data driven decision making, summarized our research and selected trends on the levels of impact and uncertainty in the following driver matrix (see Figure 7.3). First, we investigated how AI can help individuals reduce CO_2 emissions related to their behaviour. Even if the total impact is not too high compared to other interventions, it is a direct way for every person to take action. While it is probable that personalization will be used in the future, it is unclear if it will be utilized for nudging. Machine Learning models can simulate the behavioral characteristics of individuals, empowering and informing them. Furthermore, the suggested AI models can provide robust decision-making processes and climate-informed decision analysis for the collectives despite their uncertainty at global level. Their use highly depends on politics, but if they were used they could have a high impact on the global CO_2 emissions in combination with policy assessment. In our opinion, the influence of policies designed with the help of AI could be massive. There is a possibility to achieve the desired effect faster and with more support in society. It can also lead to more countries participating in large climate agreements and adhering to the targets. Unfortunately, however, there is still a great deal of uncertainty surrounding this trend, as the development of policies is not only rational but often influenced by the input of big interest groups. In addition, many factors and constraints have to be implemented that are not yet fully investigated. In addition, we presented a simulation game in

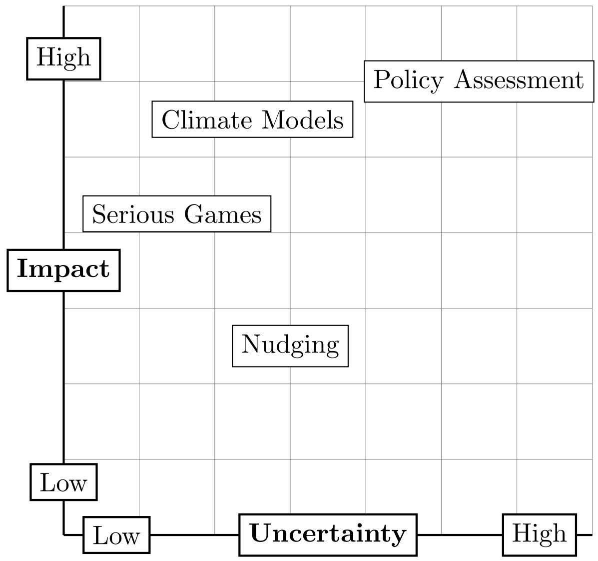

Figure 7.3: Driver matrix

a role to experience the individuals how their actions affect the global climate. Given the continuous rise in popularity, the growing number of gamers and the increasing importance of digitalization, it is rather certain that serious games will continue to shape the future of climate-related decision-making, which is why the trend is positioned rather low on the uncertainty axis. With the key ability to foster motivation and empathy towards climate issues, serious games have the potential to strongly impact people's attitudes towards climate issues and thus, their decisions, which is why the trend ranks high on the impact axis. Uncertainty and the possible risks can be a topic on the future work for the researchers. However, the research in this field provides an evidence how Machine Learning can contribute for better future decision making regarding effective adaptation and mitigation measures.

References

[1] Mario José Diván. "Data-driven decision making". In: *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)*. 2017, pp. 50–56. DOI: `10.1109/ICTUS.2017.8285973`.

[2] David Rolnick et al. *Tackling Climate Change with Machine Learning*. `https://arxiv.org/pdf/1906.05433.pdf`. [Online; accessed: 27-August-2021]. 2019. arXiv: `1906.05433 [cs.CY]`.

[3] Diana Ivanova et al. "Quantifying the potential for climate change mitigation of consumption options". en. In: *Environmental Research Letters* 15.9 (Aug. 2020). Publisher: IOP Publishing, p. 093001. ISSN: 1748-9326. DOI: 10.1088/1748-9326/ab8589. URL: https://doi.org/10.1088/1748-9326/ab8589 (visited on 08/17/2021).

[4] Daniel Kahneman. *Thinking, fast and slow.* First paperback edition. New York: Farrar, Straus and Giroux, 2013. ISBN: 978-0-374-53355-7.

[5] Cass R. Sunstein. "Nudging: A Very Short Guide". en. In: *Journal of Consumer Policy* 37.4 (Dec. 2014), pp. 583–588. ISSN: 1573-0700. DOI: 10.1007/s10603-014-9273-1. URL: https://doi.org/10.1007/s10603-014-9273-1 (visited on 07/20/2021).

[6] Emmanuel Mogaji, Sunday Olaleye, and Dandison Ukpabi. "Using AI to Personalise Emotionally Appealing Advertisement". en. In: *Digital and Social Media Marketing: Emerging Applications and Theoretical Development.* Ed. by Nripendra P. Rana et al. Advances in Theory and Practice of Emerging Markets. Cham: Springer International Publishing, 2020, pp. 137–150. ISBN: 978-3-030-24374-6. DOI: 10.1007/978-3-030-24374-6_10. URL: https://doi.org/10.1007/978-3-030-24374-6_10 (visited on 07/07/2021).

[7] Harry Brignull. *Dark Patterns: Deception vs. Honesty in UI Design.* en-US. Nov. 2011. URL: https://alistapart.com/article/dark-patterns-deception-vs-honesty-in-ui-design/ (visited on 08/01/2021).

[8] Randi Karlsen and Anders Andersen. "Recommendations with a Nudge". en. In: *Technologies* 7.2 (June 2019). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 45. DOI: 10.3390/technologies7020045. URL: https://www.mdpi.com/2227-7080/7/2/45 (visited on 06/06/2021).

[9] Hsinchun Chen, Roger H. L. Chiang, and Veda C. Storey. "Business Intelligence and Analytics: From Big Data to Big Impact". In: *MIS Quarterly* 36.4 (2012). Publisher: Management Information Systems Research Center, University of Minnesota, pp. 1165–1188. ISSN: 0276-7783. DOI: 10.2307/41703503. URL: https://www.jstor.org/stable/41703503 (visited on 07/20/2021).

[10] Min Gao, Kecheng Liu, and Zhongfu Wu. "Personalisation in web computing and informatics: Theories, techniques, applications, and future research". en. In: *Information Systems Frontiers* 12.5 (Nov. 2010), pp. 607–629. ISSN: 1572-9419. DOI: 10.1007/s10796-009-9199-3. URL: https://doi.org/10.1007/s10796-009-9199-3 (visited on 07/20/2021).

[11] Ruhi Sarikaya. "The Technology Behind Personal Digital Assistants: An overview of the system architecture and key components". In: *IEEE Signal Processing Magazine* 34.1 (Jan. 2017). Conference Name: IEEE Signal Processing Magazine, pp. 67–81. ISSN: 1558-0792. DOI: 10.1109/MSP.2016.2617341.

[12] Carlos A. Gomez-Uribe and Neil Hunt. "The Netflix Recommender System: Algorithms, Business Value, and Innovation". In: *ACM Transactions on Management Information Systems* 6.4 (Dec. 2016), 13:1–13:19. ISSN: 2158-656X. DOI: 10.1145/2843948. URL: https://doi.org/10.1145/2843948 (visited on 07/19/2021).

[13] Gourab K Patro et al. "Towards Safety and Sustainability: Designing Local Recommendations for Post-pandemic World". In: *Fourteenth ACM Conference on Recommender Systems*. RecSys '20. New York, NY, USA: Association for Computing Machinery, Sept. 2020, pp. 358–367. ISBN: 978-1-4503-7583-2. DOI: 10.1145/3383313.3412251. URL: https://doi.org/10.1145/3383313.3412251 (visited on 07/19/2021).

[14] Alain D. Starke, Martijn C. Willemsen, and Christoph Trattner. "Nudging Healthy Choices in Food Search Through Visual Attractiveness". English. In: *Frontiers in Artificial Intelligence* 0 (2021). Publisher: Frontiers. ISSN: 2624-8212. DOI: 10.3389/frai.2021.621743. URL: https://www.frontiersin.org/articles/10.3389/frai.2021.621743/full (visited on 07/20/2021).

[15] Karen Ehrhardt-Martinez, Kat A Donnelly, et al. "Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities". In: 2010.

[16] Anders Andersen, Randi Karlsen, and Weihai Yu. "Green Transportation Choices with IoT and Smart Nudging". en. In: *Handbook of Smart Cities: Software Services and Cyber Infrastructure*. Ed. by Muthucumaru Maheswaran and Elarbi Badidi. Cham: Springer International Publishing, 2018, pp. 331–354. ISBN: 978-3-319-97271-8. DOI: 10.1007/978-3-319-97271-8_13. URL: https://doi.org/10.1007/978-3-319-97271-8_13 (visited on 06/06/2021).

[17] *CycleAI: Empowering cyclists in fighting for their own safety through AI https://digieduhack.com/en/solutions/cycleai-empowering-cyclists-in-fighting-for-their-own-safety-through-ai*. en-GB. URL: https://digieduhack.com/en/solutions/cycleai-empowering-cyclists-in-fighting-for-their-own-safety-through-ai (visited on 07/19/2021).

[18] H. M. Abdul Aziz et al. "Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: a random parameter model using New York City commuter data". en. In: *Transportation* 45.5 (Sept. 2018), pp. 1207–1229. ISSN: 1572-9435. DOI: 10.1007/s11116-017-9760-8. URL: https://doi.org/10.1007/s11116-017-9760-8 (visited on 07/19/2021).

[19] Andreas T. Schmidt and Bart Engelen. "The ethics of nudging: An overview". en. In: *Philosophy Compass* 15.4 (2020). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/phc3.12658, e12658. ISSN: 1747-9991. DOI: 10.1111/phc3.12658. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/phc3.12658 (visited on 08/17/2021).

[20] David Hagmann, Emily H. Ho, and George Loewenstein. "Nudging out support for a carbon tax". en. In: *Nature Climate Change* 9.6 (June 2019). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate-change policy;Decision making;Politics;Psychology and behaviour Subject_term_id: climate-change-policy;decision-making;politics;psychology-and-behaviour, pp. 484–489. ISSN: 1758-6798. DOI: `10.1038/s41558-019-0474-0`. URL: `https://www.nature.com/articles/s41558-019-0474-0` (visited on 07/19/2021).

[21] John Weyant. "Some Contributions of Integrated Assessment Models of Global Climate Change". In: *Review of Environmental Economics and Policy* 11.1 (2017), pp. 115–137. DOI: `10.1093/reep/rew018`.

[22] Anthony Bonen, Willi Sr, and Stephan Klasen. "Economic Damages from Climate Change: A Review of Modeling Approaches". In: (Mar. 2014).

[23] Svenja Falk et al. "Politikberatung – eine Einführung". de. In: *Handbuch Politikberatung*. Ed. by Svenja Falk et al. Wiesbaden: Springer Fachmedien, 2019, pp. 3–24. ISBN: 978-3-658-03483-2. DOI: `10.1007/978-3-658-03483-2_1`. URL: `https://doi.org/10.1007/978-3-658-03483-2_1` (visited on 07/19/2021).

[24] Andreas Blätte. "Politikberatung aus sozialwissenschaftlicher Perspektive". de. In: *Handbuch Politikberatung*. Ed. by Svenja Falk et al. Wiesbaden: Springer Fachmedien, 2019, pp. 25–38. ISBN: 978-3-658-03483-2. DOI: `10.1007/978-3-658-03483-2_3`. URL: `https://doi.org/10.1007/978-3-658-03483-2_3` (visited on 07/19/2021).

[25] Alexander Graser. "Politikberatung aus juristischer Sicht". de. In: *Handbuch Politikberatung*. Ed. by Svenja Falk et al. Wiesbaden: Springer Fachmedien, 2019, pp. 39–50. ISBN: 978-3-658-03483-2. DOI: `10.1007/978-3-658-03483-2_5`. URL: `https://doi.org/10.1007/978-3-658-03483-2_5` (visited on 07/19/2021).

[26] Thomas Sterner et al. "Policy design for the Anthropocene". English. In: *Nature Sustainability* 2.1 (2019), pp. 14–21. ISSN: 2398-9629. DOI: `10.1038/s41893-018-0194-x`.

[27] Fraser J. Morgan and Adam J. Daigneault. "Estimating Impacts of Climate Change Policy on Land Use: An Agent-Based Modelling Approach". In: *Plos One* 10.5 (2015). DOI: `10.1371/journal.pone.0127317`.

[28] Stephan Zheng et al. *The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies*. 2020. arXiv: `2004.13332 [econ.GN]`.

[29] Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. "Machine Learning for Combinatorial Optimization: a Methodological Tour d'Horizon". In: *CoRR* abs/1811.06128 (2018). arXiv: `1811.06128`. URL: `http://arxiv.org/abs/1811.06128`.

[30] Richard H Moss et al. “The Next Generation of Scenarios for Climate Change Research and Assessment”. In: *Nature, 463(7282):747-756* 463.7282 (Aug. 2010). DOI: 10.1038/nature08823. URL: https://www.osti.gov/biblio/991072.

[31] Chathika Gunaratne and Ivan Garibay. “Evolutionary model discovery of causal factors behind the socio-agricultural behavior of the Ancestral Pueblo”. In: *Plos One* 15.12 (2020). DOI: 10.1371/journal.pone.0239922.

[32] Haifeng Zhang et al. “Data-driven agent-based modeling, with application to rooftop solar adoption”. In: *Autonomous Agents and Multi-Agent Systems* 30.6 (2016), pp. 1023–1049. DOI: 10.1007/s10458-016-9326-8.

[33] Janez Sušnik et al. “Multi-Stakeholder Development of a Serious Game to Explore the Water-Energy-Food-Land-Climate Nexus: The SIM4NEXUS Approach”. In: *Water* 10.2 (2018). DOI: 10.3390/w10020139.

[34] Jason S. Wu and Joey J. Lee. “Climate change games as tools for education and engagement”. In: *Nature Climate Change* 5 (2015). DOI: 10.1038/nclimate2566.

[35] Minhua Ma et al., eds. *Serious Games: Joint International Conference*. Springer International Publishing AG, 2020.

[36] Alexandra Luccioni et al. “Using Artificial Intelligence to Visualize the Impacts of Climate Change”. In: *IEEE computer graphics and applications* 41.1 (2021), pp. 8–14. DOI: 10.1109/MCG.2020.3025425.

[37] Fotis Liarokapis et al. “Multimodal Serious Games Technologies for Cultural Heritage”. In: Springer, Cham, 2017, pp. 371–392. DOI: 10.1007/978-3-319-49607-8_15.

[38] Luis Salvador-Ullauri et al. “Combined method for evaluating accessibility in serious games”. In: *Applied sciences* 10.18 (2020). DOI: 10.3390/app10186324.

[39] United Nations. *Mission 1.5*. 2020. URL: https://unric.org/en/mission-1-5/ (visited on 08/30/2021).

[40] NASA. *Climate Kids*. 2021. URL: https://climatekids.nasa.gov/ (visited on 08/30/2021).

[41] Jessica Clement. *Number of video gamers worldwide 2015-2023*. 2021.

[42] Press and Information Office of the Federal Government. *The Digital Strategy of the German government*. 2021.

[43] “Serious Gaming for Climate Adaptation—Assessing the Potential and Challenges of a Digital Serious Game for Urban Climate Adaptation”. In: *Sustainability (Basel, Switzerland)* 12.5 (2020), pp. 1–18.

[44] Jasper N. Meya and Klaus Eisenack. “Effectiveness of gaming for communicating and teaching climate change”. In: *Climatic change* 149.3-4 (2018), pp. 319–333.

[45] Stephen R. J. Sheppard et al. "Can Visualisation Save the World? - Lessons for Landscape Architects from Visualizing Local Climate Change". In: *Conference Proceedings, Digital Design in Landscape Architecture, 9th International Conf* (2008).

Chapter 8

Data Science for Animals

ELSER, SARAH
KHOMICH, MARIA

Abstract

This trend report is analyzing the state of art research and application of Data science animals and biodiversity. With the growing influence of climate change and human impact on other species, the role and necessity of such research are expanding. The well-being of animals can be classified and forecasted through observing their daily routines and behavior, and consequently drawing conclusions about their state of health and ecology. Research, on the intersection of the Ecology, Data science and Electrical engineering domains is often addressing fundamental questions in animal behaviour and becomes increasingly valuable in applied conservation settings, like monitoring populations of endangered or data-deficient species, or monitoring illegal activities in high-risk areas.

Meanwhile, the decline of insects and the resulting decrease in biodiversity is one of the major problems of our time reflected in a looming shortage of food. At the same time, there is growing interest in the influence of extreme weather conditions on biodiversity. The calculation of accurate carbon cycles is also of increasing interest, as animals have a greater influence on $CO2$ storage and release than previously thought. The object of study of current research is different groups of animals, excluding livestock, lab animals used for research, or pets. First multiple databases as well as Cornell University Archive were researched with keywords "machine learning", "artificial intelligence", "animal, biodiversity", "conservation", further speakers and their publishing from academia from the conference "AI for the planet". In the next stages, research topics were clustered into groups. The most popular groups are then identified into the following research trends: interdisciplinary research, seasonal changes research, and the role of animals in the carbon cycle.

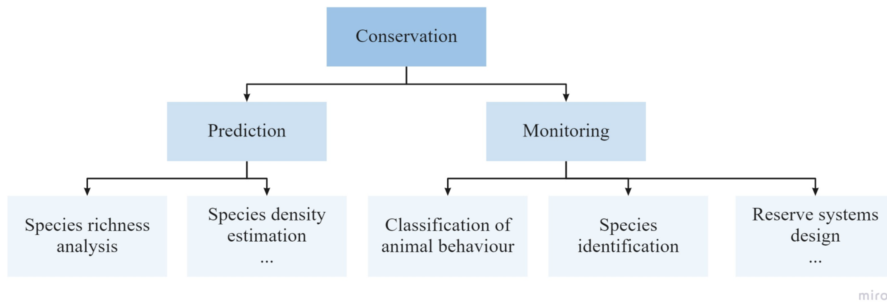

Figure 8.1: Overview of the main research directions from the analysed papers in Data science for animals.

8.1 Introduction

Following the EU Green Deal in March 2020, the EU Biodiversity Strategy 2030 sets out several actions to address the main drivers of biodiversity loss from 2021 [1].

These goals are also, from our observations aligned with the current development in the research directions:

A necessary basis for this approach are improvements in knowledge, financing and investment, which are of course interrelated, and a greater respect for nature in public and corporate decision-making.

The challenges we face today are complex and often concern the interaction between humans and their environment. When it comes to Data science for animals, it concerns mainly their interactions within ecosystems, but even these have changed through centuries of human intervention and are largely unstudied [2]. The biodiversity has severely decreased over the last years and the number of endangered species is rising. Climate change is only one of the drivers among others like degradation of habitat, over-exploitation of soils, introduction of non-native species and general pollution [3].

Current study considers only papers published after 2010 to have an overview of the most recent developments of the field. Further, the papers are clustered based on the main objective. The most popular groups are then identified into the following trends: Interdisciplinary research, seasonal changes research and the role of animals in the carbon cycle.

8.2 Trends

There are three different trends presented in the current analysis. On the one hand, within research, as the increasingly complex topics make interdisciplinary work indispensable, larger collaborations are made possible by more funding and more inter-dependencies and synergies are explored. On the other hand, seasonal fluctuations and extreme weather conditions continue to increase. The changing seasons and the associated temperatures also change the behaviour and occurrence of animals. While some animals manage to adapt to the changing conditions, others do not. Insects, for example, are very strongly affected and play a key role in our global system, so their protection is also in the interest of humans [4]. The third trend is the increased research on the contribution of animals to CO2 release and storage. Various aspects such as nutrition, habitat and range of movement play a role here.

8.2.1 Interdisciplinary research

Interdisciplinary collaboration and civic participation in biodiversity projects are increasing. Different disciplines work with separate concepts, research methods and ideas, and the collaboration of scientists, engineers, artists and social scientists, such as the a bio-tagging project [5], or projects involving species identification, requiring specific domain knowledge can lead to great added value [6].

However, mutual respect and technical language can be a challenge. Although there are more and more interdisciplinary courses of study in the university world, these influences are only slowly becoming established and a common approach must first be worked out in many areas. These programs are needed to train truly interdisciplinary scholars in the critical factor of skills and competencies [7]. Civic participation in the ecological investigation is meanwhile gaining its height, especially crowdfunding projects for species identification, taking advantage of human observational capacity in combination with Machine learning [8]. Open research sample libraries with generic recognizers also pawing the way towards more accessible and green technology [9].

Facts

- The tendency towards interdisciplinarity in the social sciences, natural sciences and engineering is increasing [10].
- Interdisciplinary papers get cited more and have more impact after approximately ten years [10].
- There are more papers currently published which mention "interdisciplinary" in their title [10].
- Not all natural sciences are equally suitable for interdisciplinary projects. Biology, and more specifically behavioural studies, is one of them. [10].

Key Drivers

- Humans feel a connection to nature, and therefore perceive it as something worth protecting and loving. This relationship can be illustrated with

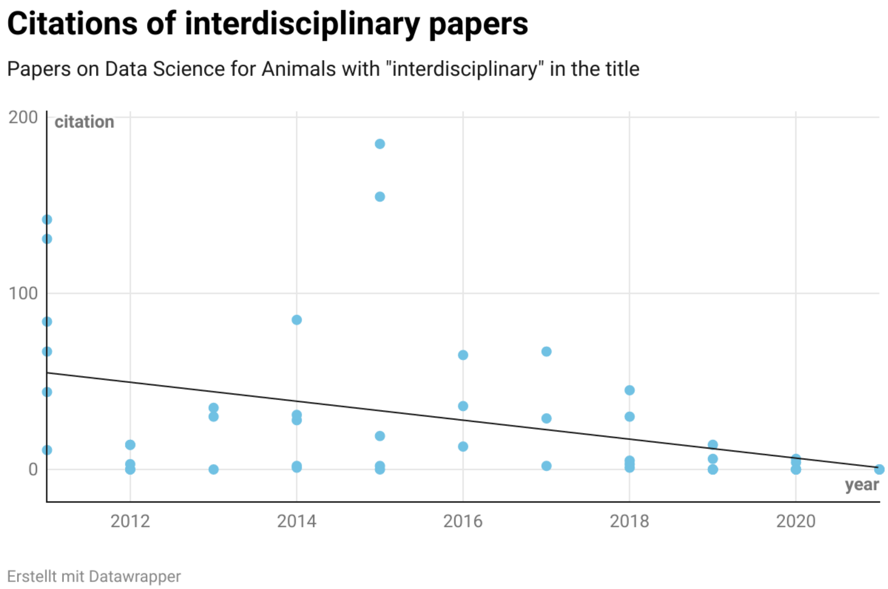

Figure 8.2: Overview of citations of interdisciplinary work on the topic of Data science for animals. Papers with "interdisciplinary" in the title get cited less in the first few years but after several years they get cited more in comparison to other papers.

different concepts and describes the innate need to be close to nature, plants and animals [11].

- In order to address global scientific challenges, the joint participation of researchers from different disciplines is necessary. These range from social sciences to life sciences to newer research fields on environment and climate [7].
- It is much easier to collect and store data in a way that they are qualitatively similar and can be read quantitatively to gain new insights. Open access libraries offer this data to everyone and make research more accessible.
- Concepts such as the Methodology for Interdisciplinary Research (MIR) have been developed to support interdisciplinary projects. MIR is supposed to give an idea how to overcome challenges and help with interdisciplinary research and communication. It provides guidance for designing an interdisciplinary education program or for approaching and outlining a research project [7].
- Social media offers opportunities for spatial, content and network analysis and are therefore increasingly used for science. This is also true for natural

science, which uses many social media platforms that offer API access for non-commercial purposes [12].

"Most interdisciplinary research centers have a tendency to become a nexus of loosely connected individuals searching for intersections, as opposed to cohesive groups tackling well-defined problems" [13].

Challenges

- Interdisciplinary projects are more expensive, they take longer, need more flexibility and are larger in size [14]. Due to the increased duration and the sometimes very particular specialization of the researchers, interdisciplinary projects are expensive. Thus, finding donors is not eternally easy and crucial projects have to be modified or do not even come to fruition. An example to this would be the African lion, a wild mammal that has been researched from many different disciplines during the last fifty years. It took about twelve years to obtain an adequate sample size for behavioral research of these animals [15].

- Highly specialized people are required to evaluate these projects, but also mediators with a broad understanding of various disciplines can be beneficial [7]. This makes it a challenge to form functioning teams that cover all essential disciplines. Within the project, it can also happen that some areas turn out to be redundant and other research results have to be included.

- Social sciences and natural sciences invoke different paradigms and epistemologies. It is essential to distinguish whether theories are based on logical empiricism or on constructivism. These differences lead to very different concepts and ways of working, and above all to different verifiability of findings [7]. Misconception or prejudices about another discipline can be a great challenge [16].

- The skills and competencies of the scientists involved have a major impact on whether the interdisciplinary project can bear fruit, and also how the projects are designed and organised [7].

- The institutional context of the research also influences the success of the collaboration [7]. Some organisations, for example the Association des États Généraux des Étudiants de l'Europe, took interdisciplinary work in account and chose their structures within their organisations accordingly to support and enhance interdisciplinarity in higher education, like increased research and successful study degrees [17].

- Despite a rapid technology integration, there prevails a lack of best-practice guidelines for researchers hoping to deploy certain technologies in the field to address particular questions [18].

- Due to the newness of the environmental research field there are not many journals with a good scientific reputation that are well respected among universities and scientists [19].

All these challenges make this research very dependent on the general interest within the field and make the research a supplicant rather than a pathfinder.

Impact

The climate and ecosystems as parts of our world are difficult to describe from a single perspective. Interdisciplinarity allows recognising new inter-dependencies and complex relationships that provide new insights. Work on complex issues is becoming increasingly interdisciplinary and demands active participation of diverse research stakeholders.

Overall, the possibilities in this area have expanded enormously through new spatial technologies like remote-sensing satellites which leads to advances in various disciplines. These technologies make it for example possible to track and analyse animals' movement and behaviour which leads to new interdisciplinary research fields like animal bio-telemetry and movement technology [2]. Above all, data collection and deployment of technologies are at the forefront of research. Especially given climate change, it is essential to work together across disciplines to find solutions to disasters that occur, to protect lives and to minimise or, if possible, prevent their impact.

8.2.2 Seasonal changes research

The far-reaching changes in climate and weather conditions are not yet actively targeted in the research, but growing interest in it is expected [20]. Seasonal changes include effects of climate change on the seasonal deviations compared to historical data and land-use change on wild population [21] and affect ecosystems at multiple scales.

Facts

- The seasonal timing of recurring biological events is a fundamental element of relationships in ecology, as well as an essential indicator of species responses to climate change [20].

- Differential shifts among interacting organisms could drive population declines through reduced productivity, increased predation or competition [20].

- Significant weather variability in the near term shows the importance of managing risks to food security [22].

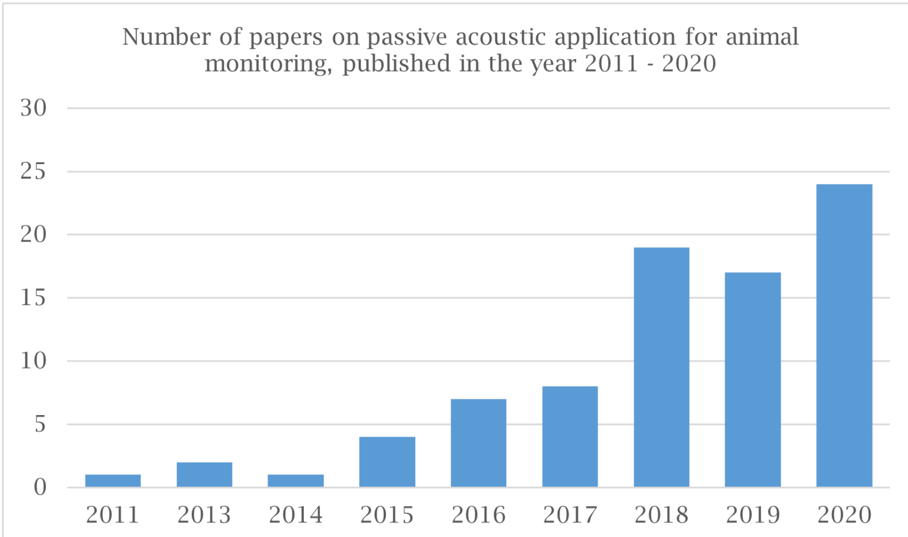

Figure 8.3: Number of papers, published in passive acoustic applications for animals and biodiversity research. 84 papers were selected randomly, based on the search with keyword combinations: "passive acoustic", "environment", "animal monitoring", "ecology" for years from 2010-2020.

Key Drivers

- Considerations of human health as extreme weather patterns create an ecological and social environment in which many infectious diseases of humans and other animals occur [23].

- Considerations of economical and natural disasters as extreme weather events may destroy habitats and food systems, forcing animals to migrate to new geographical locations, transmitting dangerous organisms [24].

- Availability of cheaper sensory network elements for collecting the spatial-temporal data at scales, including visual sensors recording images of animals that move across their field of view [21], as well as automated animal call recognition from the environmental recordings [9].

- Availability of smaller monitoring devices with GPS, time-lapse cameras and small control units to attract, classify and collect seasonal data about smaller insects [25], [26].

- The recent boom in passive acoustic monitoring (bioacoustic) sensors has offered an effective, non-intrusive and taxonomically broad means of studying wildlife populations, bio variation and biodiversity [27], [28]. Although Passive acoustic sensors were first utilised underwater during

World War I [18], since the millennium, their utility for ecological applications is employed to maintain a vast amount of analysis, including occupancy analysis, temporal abundance trends and activity patterns.

Challenges

- Collecting data on human-nature interaction is time-consuming and requires more resources than are normally obtainable.

- Tracking and monitoring behavioral changes of certain endangered animals is challenging to study and most monitoring methods are labour intensive and have a low efficiency [26].

- Certain technological challenges are also taking place, including typical challenges faced by camera trapping data, including cases where the animal [21].

Impact

Seasonal changes research, as well prediction of climate and weather-related disasters, helps to predict certain economic parameters, as well as demand in certain locations for certain products and fertilizers. Investigating ecological correlates of diversity and indicators of diversity in a single geographic area provides a better understanding of general patterns across groups. Different environmental conditions, especially precipitation [29], predict well the structure of species richness and diversity.

Currently, for example, agriculture and logging are pervasive in the tropics hunting and trapping is the most geographically widespread threat to mammals and birds [30], research in this area can help to craft better risk mitigation strategies for different scenarios.

As some of the use cases for Data science for animals might not require state-of-the-art model accuracy but effective and intelligent process management, the increased usage of small devices is of particular interest. Such devices allow the model to be fast enough to process examples in real-time. Using Machine Learning on small devices – Internet of Things (IoT) devices, for example, - could benefit from different perspectives. Not only can it collect a big amount of data, but also have sufficient power to analyze these collected data and extract useful information from huge data sets. Moreover, it has the advantage of latency, as the data can be processed directly on the device. Another valuable addition here could be data privacy since the local data again will be processed locally. Although the model, which could work on the IoT device should be fast, more can be processed. And since all the work is done on the device the most power-hungry process - communication - is not needed.

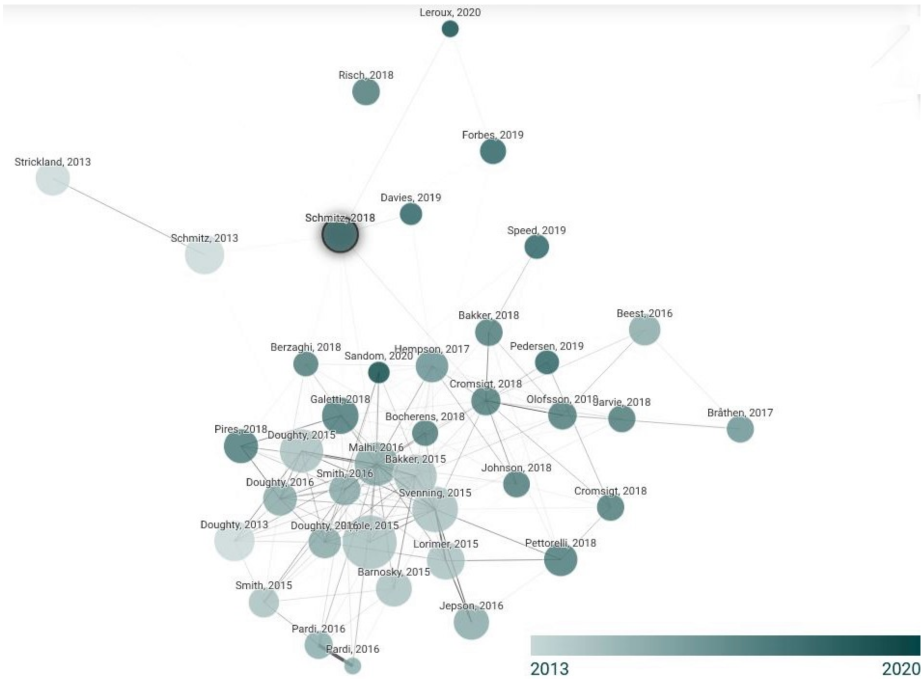

Figure 8.4: Research of animals and the carbon cycles are very current and originate from research of rewilding and conservation.

8.2.3 Role of animals in the carbon cycle

Wild animals have a bigger influence on carbon cycles than previously thought. Many are familiar with the Methane production by cows [31] but they are not the only species that influence the climate. Animals move huge amounts of carbon through their migration and feeding habits, and some key species have a major impact on ecosystems. With the increase in technical capabilities and Data science, these connections can be better explored. This can also be observed in the graph below. With the addition of technology, new fields of research have formed, as explained above, from fields that have existed for some time. For example, the impact of animals can also be brought into the carbon cycle.

Facts

- Animals such as lions or elephants roam widely across landscapes, creating spatial dynamics that affect soil mixing and determine and change soil carbon storage [32].

- Animals can influence biophysical conditions such as temperature by destroying vegetation or trampling [33].

- The food webs form complex top-down and bottom-up relationships between producers, consumers, and decomposers. The interdependencies between these three types are rather complex and their interactions are fundamentally nonlinear [34]. For example, the multi- and interdisciplinary interactions between prey and predators are a key mechanism for explaining the evolution of organisms in their ecosystems, controlling ecosystem functions like energy flow and the biogeochemical cycle [3].

- Animals influence their ecosystem also through their eating habits. Insects play a key role in pollinating about 80 per cent of the world's plants, contributing to diversity and CO2 conversion [35]. Frugivores for example help with the biodiversity of plants when they excrete plant seeds. Herbivores feed on plants that otherwise bind Co2 with the help of photosynthesis [36]. Predators interact with herbivores and can cause a reverse effect. These eating patterns of animals influence the spatial patterning of vegetation which in return has an impact on the biomass and carbon uptake of their surroundings [37].

Key Drivers

- To meet the challenges of climate change accurate calculations of biomass carbon should be done. They are essential to extend existing models of the carbon cycle to test their accuracy, as play a key role in predicting global changes in biogeochemical functions [32].

- Together with the seasonal changes research trend, this trend key drivers include quantitative insights with new technologies like, biotagging, computer vision models that can recognize animal patterns [38], and remote sensing make it possible to follow animals from a far and understand landscapes better.

- Dynamic food-web models can be calculated to portray prey predator interactions and to understand animal behaviour better [3].

- Reports such as the IPCC report [39] highlight the importance of this. As the EU and other countries are willing to invest much more money in scientific projects, the pressure for knowledge and the economic value in these projects and questions is high.

- The reliance of food on agriculture, which is significantly influenced by animals and plants diseases [40].

- The influence of the civil participation trend, allowing spatial and temporal perspective analysis to reveal hotspots and areas with pressure on biodiversity [12], [8].

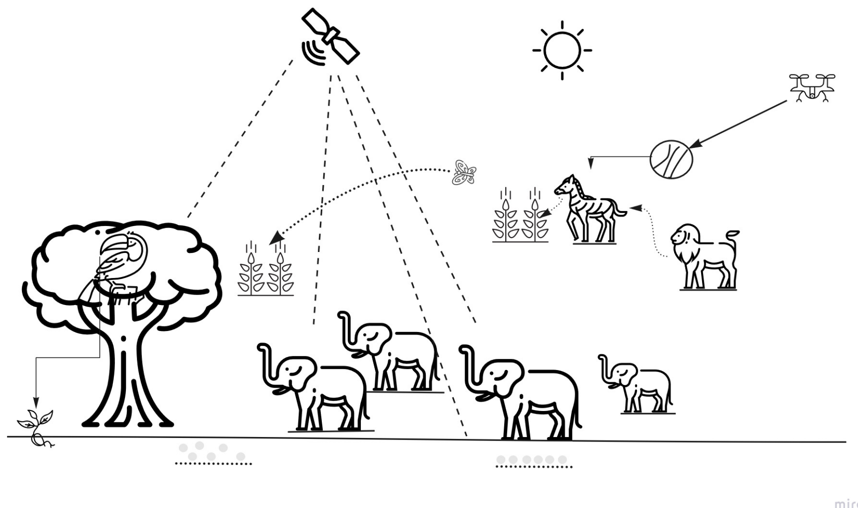

Figure 8.5: An overview of interactions and interdependencies of animals and how technology is able to track them.

Challenges

- There is a lack of data regarding animal's behaviour, topography, habitat structures, spatial arrangement of habitat patches and habitat connectivity within and among ecosystems across landscapes [41].
- Until know there is only limited ability to detect and quantify animal biomass and animal effects [42].
- Many interdependencies due to food chain and behaviorism are unclear [3].

Impact

The deeper understanding of the roles that animals play within their ecosystem also opens up the possibility to calculate carbon cycle models more accurately. With the latest IPCC report [39] estimating that 1.5 degrees will be reached by 2030, it is even more fundamental to change our interactions with the environment and wildlife. Animals not only suffer from climate change, they also play a crucial role in its ecosystems and carbon cycles. In addition, there are many other impacts on our society, including changing landscapes and land use, issues of biodiversity loss, deteriorating food safety and the use of chemicals [40]. A single animal may not trigger an entire typhoon with a single wing movement, but its contribution to our climate should not be overlooked.

8.3 Conclusion

Data science has an enormous opportunity to support animal conservation and biodiversity research and some positive shifts are presented in this area. Few researchers have addressed the problem of seasonal changes and their influence on animal health and migration behaviour, giving both: high impact, as well medium uncertainly. Two other trends, presented in the current analysis are closely coupled with each other, as they all are influenced by global environmental and technological trends, such as huge volumes of accessible data [43], globalization, connectivity and knowledge culture [44].

Animals not only suffer from climate change, but they also play important roles in their ecosystems and carbon cycles. At the moment these roles are not researched well enough to simulate accurate carbon cycle models. A single animal may not immediately trigger an entire typhon with one movement of its wings, but their contribution to our climate should not be overlooked.

Pop culture is actively working on public involvement in biodiversity topics in order to shape consumer behavior and consciousness by movies like – "Seaspiracy" [45], "Cowspiracy". These movies, while shocking to the general public, also triggered a lot of discussions in scientific communities – from discussions on food sustainability [46].

We see how pop culture is affecting research, and it is disappointing to witness, however, how low is the influence of the research on business. While risks from biodiversity loss are looming, less than one quarter of "at risk" companies worldwide report on their influence on biodiversity, with mining being the only "at risk" sector in which a majority of companies report risks from biodiversity loss [**kpmg2020time**].

Currently, only 5 percent of Climate Tech start-ups in the EU are working on the biodiversity topic [47]. While this percentage is incredibly small, there is also an opportunity. With the other sectors of climate deal being more represented by already big and competitive tech start-ups and players. biodiversity together with sustainable finance is still an open niche. Meanwhile, a pool of capital is increasingly flowing to environment-focused companies creating even more incentive to develop in this direction.

References

[1] The European Commission. *EU climate action and the European Green Deal. 2030 climate and energy framework.* The European Commission, 2021. URL: `https://ec.europa.eu/clima/policies/strategies/2030_en`.

[2] James Cheshire and Oliver Uberti. *Where the animals go: tracking wildlife with technology in 50 maps and graphics.* Particular Books London, 2016.

[3] Florian D Schneider et al. "Animal diversity and ecosystem functioning in dynamic food webs". In: *Nature Communications* 7.1 (2016), pp. 1–8.

[4] Christopher A Halsch et al. "Insects and recent climate change". In: *Proceedings of the national academy of sciences* 118.2 (2021).

[5] Christian Nold et al. "Biotagging Manchester: Interdisciplinary Exploration of Biodiversity". In: *Leonardo* 44.1 (2011), pp. 66–67.

[6] Jaime R Ticay-Rivas et al. "Spider specie identification and verification based on pattern recognition of it cobweb". In: *Expert systems with applications* 40.10 (2013), pp. 4213–4225.

[7] Hilde Tobi and Jarl K Kampen. "Research design: the methodology for interdisciplinary research framework". In: *Quality and quantity* 52.3 (2018), pp. 1209–1225.

[8] Steve Kelling et al. "ebird: A human/computer learning network for biodiversity conservation and research". In: *Twenty-Fourth IAAI Conference*. 2012.

[9] Michael Towsey et al. "A toolbox for animal call recognition". In: *Bioacoustics* 21.2 (2012), pp. 107–125.

[10] Richard Van Noorden et al. "Interdisciplinary research by the numbers". In: *Nature* 525.7569 (2015), pp. 306–307.

[11] Uta Eser et al. "Prudence, justice and the good life: a typology of ethical reasoning in selected European national biodiversity strategies". In: *Bundesamt für Naturschutz, Bonn* (2014).

[12] Tuuli Toivonen et al. "Social media data for conservation science: A methodological overview". In: *Biological Conservation* 233 (2019), pp. 298–315.

[13] Diana Rhoten. "Interdisciplinary research: Trend or transition". In: *Items and Issues* 5.1-2 (2004), pp. 6–11.

[14] Catherine Lyall et al. "Short guide to reviewing interdisciplinary research proposals". In: *ISTTI Briefing Note* 2 (2007).

[15] Craig Packer. "The African lion: a long history of interdisciplinary research". In: *Frontiers in Ecology and Evolution* 7 (2019), p. 259.

[16] Michael Redclift. "Dances with wolves? Interdisciplinary research on the global environment". In: *Global Environmental Change* 8.3 (1998), pp. 177–182.

[17] W James Jacob. "Interdisciplinary trends in higher education". In: *Palgrave communications* 1.1 (2015), pp. 1–5.

[18] Ella Browning et al. *Passive acoustic monitoring in ecology and conservation*. WWF Conservation Technology Series, 2017.

[19] Lisa M Campbell. "Overcoming obstacles to interdisciplinary research". In: *Conservation biology* 19.2 (2005), pp. 574–577.

[20] Sarah R Weiskopf et al. "Climate change effects on biodiversity, ecosystems, ecosystem services, and natural resource management in the United States". In: *Science of the Total Environment* 733 (2020), p. 137782.

[21] Xiaoyuan Yu et al. "Automated identification of animal species in camera trap images". In: *EURASIP Journal on Image and Video Processing* 2013.1 (2013), pp. 1–10.

[22] Vincent Gitz et al. "Climate change and food security: risks and responses". In: *Food and Agriculture Organization of the United Nations (FAO) Report* 110 (2016).

[23] Anthony J McMichael. "Extreme weather events and infectious disease outbreaks". In: *Virulence* 6.6 (2015), pp. 543–547.

[24] Matt McGrath. *Deadly olive tree disease across Europe 'could cost billions'*. `https://www.bbc.com/news/science-environment-52234561`. Accessed: 2021-08-27. 2020.

[25] Ella Browning et al. "Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds". In: *Methods in Ecology and Evolution* 9 (3 2018), pp. 681–692.

[26] Toke T Høye et al. "Deep learning and computer vision will transform entomology". In: *Proceedings of the National Academy of Sciences* 118.2 (2021).

[27] Rory Gibb et al. "Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring". In: *Methods in Ecology and Evolution* 10 (2 2019), pp. 169–185.

[28] Czesław Klocek Weronika Penar Angelika Magiera. "Applications of bioacoustics in animal ecology". In: *Ecological Complexity* 43 (2020), pp. 30–059.

[29] Andrea Paz et al. "Environmental correlates of taxonomic and phylogenetic diversity in the Atlantic Forest". In: *Journal of Biogeography* (2021).

[30] Michael BJ Harfoot et al. "Using the IUCN Red List to map threats to terrestrial vertebrates at global scale". In: (2021).

[31] EH Cabezas-Garcia et al. "Between-cow variation in digestion and rumen fermentation variables associated with methane production". In: *Journal of Dairy Science* 100.6 (2017), pp. 4409–4424.

[32] Oswald J Schmitz et al. "Animals and the zoogeochemistry of the carbon cycle". In: *Science* 362.6419 (2018).

[33] Oswald J Schmitz et al. "Animating the carbon cycle". In: *Ecosystems* 17.2 (2014), pp. 344–359.

[34] Pierre Legreneur, Michel Laurin, and Vincent Bels. "Predator–prey interactions paradigm: a new tool for artificial intelligence". In: *Adaptive Behavior* 20.1 (2012), pp. 3–9.

[35] Sebastian Seibold et al. "Arthropod decline in grasslands and forests is associated with landscape-level drivers". In: *Nature* 574.7780 (2019), pp. 671–674.

[36] Christopher J Sandom et al. "Trophic rewilding presents regionally specific opportunities for mitigating climate change". In: *Philosophical Transactions of the Royal Society B* 375.1794 (2020), p. 20190125.

[37] Fabio Berzaghi et al. "Assessing the role of megafauna in tropical forest ecosystems and biogeochemical cycles–the potential of vegetation models". In: *Ecography* 41.12 (2018), pp. 1934–1954.

[38] Microsoft. *Wild Me Kernel Description*. 2021. URL: `https://www.microsoft.com/en-us/ai/ai-for-earth-wild-me` (visited on 09/02/2021).

[39] IPCC. *AR6 Climate Change 2021. The Physical Science Basis)*. IPCC, 2021. URL: `https://www.ipcc.ch/report/ar6/wg1/`.

[40] Katy Wilkinson et al. *Infectious diseases of animals and plants: an interdisciplinary approach*. 2011.

[41] S Braaker et al. "Assessing habitat connectivity for ground-dwelling animals in an urban environment". In: *Ecological applications* 24.7 (2014), pp. 1583–1595.

[42] SC Stark et al. "Differences in Amazon forest growth and carbon dynamics predicted by profiles of canopy leaf area and light environment derived from LiDAR". In: *AGU Fall Meeting Abstracts*. Vol. 2011. 2011, B44C–07.

[43] Alem Tedeneke. *Futures Report Outlines Top Trends Impacting Global Economy, Society and Technology*. `https://www.weforum.org/press/2021/04/futures-report-outlines-top-trends-impacting-global-economy-society-and-technology-bdfe790c3a/`. Accessed: 2021-09-02. 2021.

[44] *Die Megatrends*. `https://www.zukunftsinstitut.de/dossier/megatrend-wissenskultur/`. Accessed: 2021-09-02.

[45] Karen McVeigh. *Seaspiracy: Netflix documentary accused of misrepresentation by participants*. `https://www.theguardian.com/environment/2021/mar/31/seaspiracy-netflix-documentary-accused-of-misrepresentation-by-participants`. Accessed: 2021-08-19. 2021.

[46] University of Gothenburg. *Event. Seaspiracy: Researchers React*. `https://www.gu.se/en/event/seaspiracy-researchers-react`. Accessed: 2021-08-27. 2021.

[47] Ines Streimelweger. *Creandum and Speedinvest Report: The Growth and Future of Climate Tech Startups in Europe*. `https://blog.creandum.com/speedinvest-creandum-report-the-growth-and-future-of-climate-tech-startups-in-europe-3beced57731f`. Accessed: 2021-08-27. 2021.

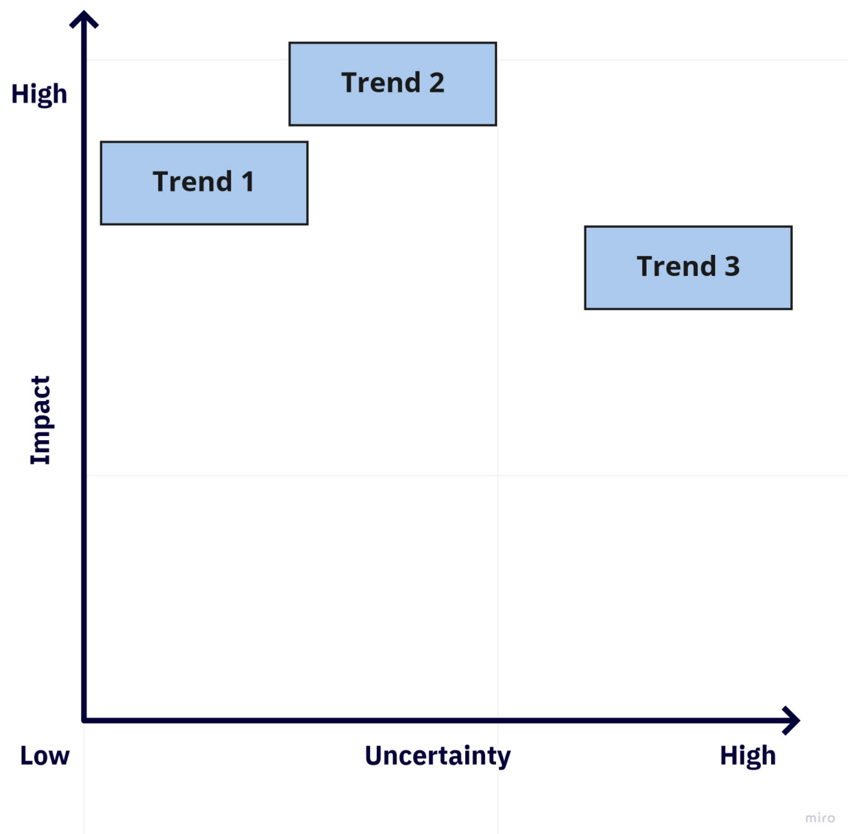

Figure 8.6: Figure 5: Trends matrix with x-axis describing the level of uncertainty of a trend and y-axis describing the impact level, where trend 1 - interdisciplinarity trend, trend 2 - seasonal changes trend, trend 3 - role of animals in the carbon cycle trend.

Chapter 9

Forestry Agriculture and Data Science

CAO, JIMIN CAO
GLIGORIC, PETAR
JAIEM, MEHDI
MANDHOUJ, MALIK
ÖZBAY, ABDULLAH CEM
SALEM, MONCEF BEN HADJ
TÜLÜ, MINE
YAN, HONGFEI
ZHANG, LINJING
ZHANG, YANG

Abstract

Agriculture and forestry are significant for the sustainability of human life on earth. Agriculture refers to protecting the global human population with crop plants that provide food and other products. Considering the increase in the world population, we need to reduce costs and increase productivity in agriculture to provide enough food for people. We need forests to survive and for many fundamental raw materials we use. Forests provide a habitat for animals and a livelihood for humans. It also prevents climate change.

Agriculture and forestry are complex businesses aiming to maximize yields by optimizing various variables. Data obtained from forests and agricultural lands through different advanced technologies will play a key role in optimizing forestry and agriculture.

The article analyzes trends in technologies for optimizing agriculture and forestry, incorporates data science to improve forestry and agriculture, and estimates future trends.

9.1 Introduction

Agriculture and forestry are key drivers of climate change. We need to optimize farming and build smart forestry to protect forests and benefit from them to tackle climate change. In recent years, the most promising approach seems to be data-driven agriculture. The aim is to collect a large amount of data from farms and forests and optimize farming and applications if forestry using obtained data. The technology-driven approach shows some significant advantages, such as saving money and efficient work, increasing production or reducing costs with the least effort, and producing quality food with more environmentally friendly practices. Moreover, current agriculture is an information-based farming system that increases farm production efficiency, productivity, and profitability in the long run.

To benefit from data science in agriculture and forestry, producers must use systems that generate data on their farms. This data can be obtained by devices such as the Insitu sensor, which supports continuous, real-time monitoring of forest conditions. While doing these, it aims to minimize the negative impacts on the environment. In addition, it prevents misuse of resources and the pollution of the domain. That will help farmers to decide strategic and operational decisions properly. Traditionally, farmers have gone to the fields to check their crops' status and make decisions by observation. This approach is no longer sustainable, as some areas are too large to be efficiently managed according to people's efficiency, sustainability, and availability.

The Internet of Things (IoT) refers to the cohesive use of sensors and other devices in an agricultural context. The IoT turns every farming-related item and action into data. IoT technologies are one of the reasons agriculture can generate such a large amount of valuable information. With the increase of obtained data and the IoT, the amount of data in agriculture increases at an extraordinary rate. Thus, agriculture has become another focus of big data research. Big data technologies can be applied to solve storage and analytic problems of a large amount of data. Big data also offers new methods and ideas for agricultural research and business development.

This report will briefly describe the latest research and studies of some technologies in smart forestry and agriculture. The article consists of four trends about the technologies used to optimize agriculture and forestry. Finally, the challenges and improvements of data science in forestry and agriculture are analyzed, and the future development trends are estimated.

9.2 Trends

While examining the impact of data science on agriculture and forestry, the most important and promising trends can vary in many different fields of study. IoT in agriculture, big data and smart farming, hardware specializations, embedded system tools are the most epoch-making approaches. In order to meet the ever-increasing demand for food, production efficiency should be increased

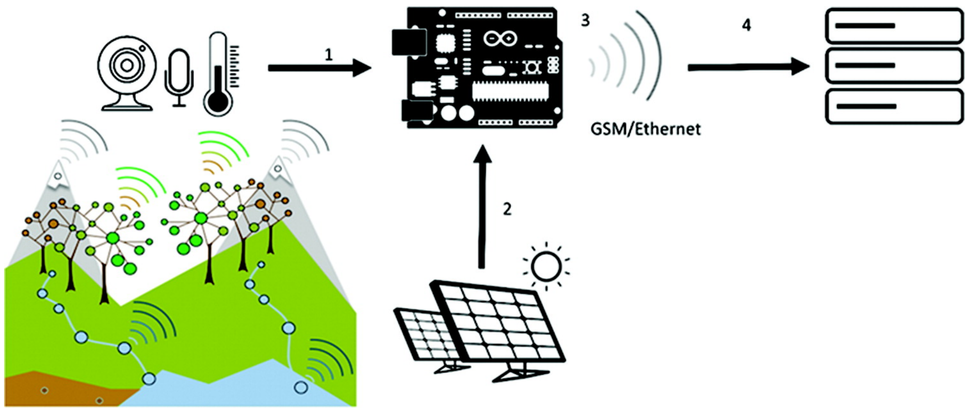

Figure 9.1: Field data collection from sensors that acquire data in microcontrollers (1) connected to, e.g., autonomous supply of energy (2) and send the data through GSM/Ethernet technologies (3) to a local server (4).

by automation and other technologies in data science. This report aims to demonstrate some of the important trends of data science in agriculture and forestry.

9.2.1 Sensor-supported forestry management

The massive increase in food demand has made forecasting necessary because of forestry's potential to store and capture carbon dioxide; sustainable forestry management will significantly mitigate climate change. This emerging branch of forestry management is what we call climate smart forestry. Of course, it relies on much higher data requirements than traditional forestry. New devices like situ sensors that support continuous monitoring of forest conditions in real-time can meet these data requirements. See the data collection and wireless data transmission example in Figure 9.1 [1].

Facts

- According to the Kyoto Protocol (KP), Carbon sequestration in terrestrial sinks can be used to offset greenhouse gas emissions. Currently, European forests absorb 7 to 12% of European emissions with agricultural land being a source and forests a sink of CO2 [2].

- In July 2016, the European Commission (EC) published a legislative proposal for incorporating greenhouse gas emissions and removals due to Land Use, Land Use Change and Forestry (LULUCF) into its 2030 Climate and Energy Framework. The Climate and Energy Framework aims at a total emission reduction of 40% by 2030 for all sectors together as part of the Paris Agreement [3]. Through Climate smart forestry EU has the potential to achieve an additional combined mitigation impact of 441 Mt CO2/year by 2050 [4].

- As technology evolves, the sensors available for forestry and agriculture are becoming more precise and cheaper, and the transmission methods are faster and more widespread [5].

- New generation of increasingly reliable sensors and data transmission and processing tools makes climate smart forestry steadily more possible. In addition, the mounting of sensors on forest machinery provide new avenues for collecting monitoring data [1].

Key Drivers

- These new sensors make data available in near real time, and multiple stakeholders can easily access the data. Older manual inspections do not provide accurate and varied data, making it difficult to achieve environmentally intelligent forestry. Sensor-supported forestry management allows people to optimize forestry management more efficiently, making it more sustainable and climate smart.

- Increasing the spatial and temporal scales of the monitoring. 3D forestry models can be built to monitor and link forestry resources from multiple perspectives, which allows quicker and more objective access to data, and restructuring of forestry to be more environmentally friendly.

- On-site sensor data is a good supplement to the remotely sensed data. Airborne and UAV (unmanned aerial vehicle) sensors have limitations that make it difficult to provide detailed information for some applications in real time.

Challenges

- The large area coverage of the Global System for Mobile Communications these days makes data transmission easier. However, the distance of the communication towers can still cause difficulties for implementation.

- Large amounts of data from sensors need to be stored and processed. It's still a common challenge in many areas.

- The Sensors can cover large areas, but The distance in large ecosystems is still a problem. How to set the number of sensors and the distance between them can significantly affect the reliability and stability of the data, There are many uncertainties in the natural world.

- Even with solar panels and lithium-ion batteries, long-term energy supply can be difficulties, for example in some forests with closed canopies.

Impact

Forests play a significant and beneficial role in mitigating climate change. At the same time, climate change has also been identified as the main driving force for changes in forest productivity. Therefore, changes in the forest can reflect climate changes to a large extent. The impacts of sensor-supported forestry management are apparent. As soon as possible, we can receive timely information on forests and take relevant measures to prevent forest disasters, such as forest fires and forest pests. Besides, forest productivity (such as Oxygen production/Carbon dioxide absorption, wood production, and so on) can be increased through forest management. In addition, we can obtain statistical data such as the number and types of vegetation in the forest to protect the diversity of vegetation easily. In summary, sensor-supported forestry management is a very significant research trend in the next few decades.

9.2.2 Embedded systems in agriculture

The massive increase in food demand has made forecasting necessary in agricultural management practices. Precision farming can significantly reduce the amount of fertilizer, herbicide, or seed rates while increasing the yield of crops. Through precision farming, real-time data on soil, temperature, and moisture are collected and processed by sensors placed throughout the farming field. This data can be analyzed to help farmers make the best choices about planting, fertilizing, and harvesting their crops. Precision livestock farming, for example, utilizes sensors, microphones, embedded systems to monitor animal activity, detect disease, improve health and welfare or production/reproduction. However, this technology requires an extensive IT infrastructure and the intervention of many professionals, so it is mainly dominated by megacorporations and increases their revenues even further. Therefore, a smaller but powerful embedded system is required. Smaller farmers or farms could use this system in areas without WiFi networks to reduce costs and increase yields.

Facts

- In 2019 farmers have saved up to 8,700 Eur per year [6] and reduced greenhouse gases by 10% by using smart farming technologies like embedded systems.
- In Europe alone, between 70% and 80% of farming equipment being sold is incorporating precision agriculture technology in some way [7].
- In order to access and predict plant growth dynamics, an RNN model using the LSTM architecture was implemented on a Raspberry Pi 3 model B and lasted for up to 180 days on battery [8].
- By creating an Ad Hoc wireless sensor network, small and low energy consuming embedded systems can easily be configured to communicate with each other over large areas of land [9].

	Laptop	Raspberry Pi
Size	Like a Pizza	Like a Cookie
Cost	over 1500 Euro with GPU	up to 150 Euro with GPU
Mobility	lower	higher
Computing time	faster	slower
Power consumption	about 25 W	about 2.5 W

Figure 9.2: Features Comparison of laptop and Raspberry Pi [10]

- A low-power embedded system can also run artificial intelligence algorithms with complex structures, such as CNNs. Embedded systems also have high mobility compared to desktop computers and laptops [10].

Key Drivers

- Embedded systems often contain a sensor module that collects relevant information such as soil humidity and temperature in real time and will autonomously upload the data via Bluetooth communication. This eliminates the need for expensive and large IT infrastructure or WiFi connection.

- The embedded system with an external Graphics Processing Unit (GPU) can collect and analyze large amounts of data in real time by using machine learning algorithms such as CNN, so it is not necessary to save the data in large storage devices. The implementation of such real-time embedded systems decreases cost and increases efficiency.

- Some embedded systems use only the Raspberry Pi for computing, which could be powered by a battery. The use of such low-power embedded systems in large-scale agricultural production can significantly reduce costs and energy consumption. Figure 9.2 shows the advantages of the Raspberry Pi compared to a normal computer or laptop: lower cost, higher mobility and lower energy consumption.

- Compared to a normal computer or laptop, the size and shape of the embedded system device are smaller and more portable. This improves the mobility of the device in a large-scale agricultural production.

Challenges

- Wireless sensor networks for data collection are currently being developed as wired systems have proven to be too expensive and cumbersome to manage in an agricultural environment [9].

- Some technological constrains still exist. Most notably communication between sensor nodes and embedded systems without interference, with low power and at large distances while still remaining cost effective [11].

- The limited RAM and cache memory of some embedded systems leads to memory overflow which negatively impacts power consumption [8].

- Privacy issues are a challenge by using embedded systems. How to protect data and prevent data leakage is a huge problem in embedded system.

- During data collection, delays at information transfer due to bad weather conditions or hardware damage can affect the process of system operation.

- Embedded systems have less computing power than desktop computers or laptops and will take longer time to run the same algorithms (possibly more than 10 times longer) [10]. The computational power of the system limits the number of possible clients.

Impact

The use of machine learning algorithms running on embedded systems, without involving the transfer of large amounts of data from cloud computing servers, is a way to help improve agricultural yields. In addition embedded systems are characterised by low energy consumption and high efficiency. Less energy consumption can help mitigate climate change. Greater portability and mobility make it possible to use such embedded systems on a large scale in agriculture, these systems can be extended and used on mobile platforms e.g. drones. This could help alleviate food crises and food shortages in the future.

9.2.3 Big data and IoT in agriculture

As the population grows, humankind will need more crops and food to maintain the circle of life. It is necessary to increase production to feed the growing world population. The digital transformation of agriculture and information technology based on big data and IoT plays a crucial role in solving this problem. Through big data, farmers enable to monitor their agricultural activities and farm conditions in real-time. Applying big data analytics and IoT agriculture can significantly benefit and use the land to create more agricultural value.

Facts

- Big data in agriculture relies on utilizing information, technology, and analytics to create valuable data that farmers can utilize to monitor agricultural activities and farm conditions. It can be studied under two significant areas, such as precision agriculture and smart farming.

- Precision Agriculture contains the implementation of automatically controlled agricultural machines, monitoring the yields [12], and various seed drilling and fertilizer spreading methods. Smart farming is a new, evolving, and developing concept. Big data applications in farming are not strictly about primary production but play a significant role in improving the efficiency of the entire supply chain and business process.

- Big data can solve problems that farmers are faced. Such decisions can include deciding on the best seed varieties or the most efficient agronomic practices or which market opportunities to pursue to ensure maximum productivity. Given the industry's growing labor shortage, big data analytics capabilities could reduce the need for manual labor, which is a massive advantage for agriculture [13].

- While increasing the efficiency of usage in resources, IoT also offers mobility and agility. In addition, it reduces the product waste caused by human-made production and thus provides better quality food and a more profitable business.

- The main steps of applying IoT technology in agriculture are collecting data via sensors and intelligent tools [14], as following transmitting the data with IoT devices to a cloud platform to analyze. Finally, farmers will examine collected data to visualize and manage the proper operations due to obtained information.

- Applications of the new technology in agriculture can be listed as greenhouse production, environment measurement, sensing local agricultural parameters, crop monitoring, animal movement monitoring, and data sharing via cloud computing [15].

- Recent reports declare that the IoT device installation will see an associate annual growth rate of 20% in the agriculture industry. And the number of agricultural connected devices will grow from 13 million in 2014 to 225 million by 2024. The global agricultural market size in 2018 is 14.79 billion dollars, while the predicted market size in 2023 is 28.64 billion dollars [16].

- The most effective and most used sensors in agriculture are electrochemical sensors that measure pH and soil nutrient levels. Optical sensors use light to measure soil properties and determines the soil's clay, organic matter, and moisture content. Mechanical sensors measure soil compaction or mechanical resistance, dielectric soil moisture sensors measure moisture levels in soil, and airflow sensors detect soil air permeability [14].

- Collected data is transmitted with different communication protocols like MQTT, TCP, and FTP, which are conveyed to the servers [17]. After the local computation, the predicted current state is provided as the agricultural services.

Key Drivers

- Big data can be solved by the improvement of software and hardware. The software should analyze the data and present the results to farmers in an accessible form. Moreover, improving the analytical capabilities of machine learning and AI systems is essential to reduce the impact of software constraints on our data analysis [18]. On the hardware side, there are a variety of sensors that collect available data. Therefore, the hardware devices need to continue to evolve to support a large amount of data information.

- UAV drones for agriculture are built to capture accurate information that airplanes and satellites can not collect. A drone construction includes propulsion and navigation systems, GPS, programmable controllers, and equipment for automated flights [19]. Drones have the functionality of crop monitoring, crop spraying, fighting infections, and seeding the deforested areas.

- An IoT sensor can be utilized, which gives real-time information about the insects on plants and the health of the crops. These sensors are able to capture images of pests the naked eye can not see [20].

- With data collection, crop storage will have new storage management because the data consists of information from last year's products and storage capability [21]. Classifying the data provides insurance to farmers, while IoT devices support better farm management.

Challenges

- Collecting and managing big data is expensive due to the amount and variability of data. On the one hand, algorithms that mine small amounts of data cannot be directly applied to big data. Big data in agriculture has its particularity. In many cases, algorithms need to strike a balance between timeliness and accuracy of processing.

- The obstacles of IoT in agriculture can be listed such as the massive data analysis, cost of expensive hardware components, poor internet connectivity, which may cause complexity in IoT applications [22].

- No field or paddocks, even the plants at your home, are entirely the same; the obtained characteristics of water, soil, and air quality may alter in distinct locations.

Impact

Big Data and IoT are two essential technologies for the development of agriculture. As we all know, agriculture is vital to people, and if we can monitor crops accurately with minimal labor, we can increase crop yields, which is a benefit to the world. Moreover, it will also boost the market economy. In the past, farming was achieved with traditional techniques, which include minor technological applications. In the future, more complete farm services will be offered by IoT devices and big data, which yield to analyze soil, crop, weed, and pest variables thanks to customized applications suitable for agriculture and devices communicating with each other [23]. As farm productivity improves, it also provides valuable feedback for agricultural decisions. Thus a farm's quality can be noticeably improved. Obtained big data also encourages the analysis of different aspects of information, which is easier to consider essential keys for agriculture.

9.2.4 Remote Sensing for wild land fire crisis management

Introduction

Recent megafires in Greece, Australia, Brazil, Indonesia, and The United states have not only devastated ecosystems but also contributed to climate change via carbon emissions [24]. Thus, finding better ways to deal with this major environmental issue is of great importance. Recent advancements in machine learning, computer vision, and remote sensing technologies provide new tools for detecting and monitoring forest fires, while novel materials and microelectronics have allowed sensors to be more efficient in recognizing active forest fires.

Key drivers

- Remote sensing technologies provide continuous, dynamic and large-scale observational data, which further promotes the research and development of forestry technologies [26].

- Remote sensing is a cost-effective approach to collect data per unit area. Although the financial expenses of remote sensing can be high at times—for example, building, launching, and operating satellite remote sensing systems is costly, making certain imagery expensive—much of it is freely available. Furthermore, while commercial remote-sensing devices may appear to be pricey, the data still allows for low-cost assessment on a unit-area basis [25].

- A new generation of satellite sensors, combined with unmanned aerial vehicle (UAV) technology, has resulted in a superior synergy of current and future remote sensing technology, allowing for better monitoring of the extent and frequency of forest fires [27].

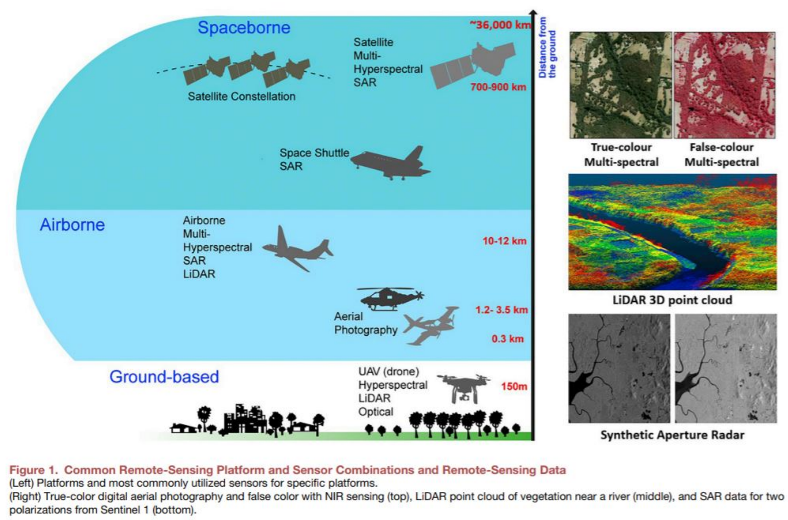

Figure 9.3: Common Remote-Sensing Platform and Sensor Combinations and Remote-Sensing Data [25]

- Nanomaterials and microelectronics advancements have lately enabled the use of tiny low-Earth-orbiting satellites known as CubeSats. In terms of smoke and fire detection, CubeSats have major advantages over traditional satellites, as they are more cost-effective and have better temporal resolution/response time [28].

Facts

- Wildfires ravaged 150 to 250 million hectares of tropical forests every year. Human activities are responsible for 90% to 95% of all such fires [29].

- According to studies on the environmental implications of tropical wildfires, biomass burning results in substantial carbon emissions, large amounts of trace gases and aerosol particles, and black carbon release of over 100 million tons of smoke aerosols into the atmosphere [29].

- While forest fires can contribute to climate change through carbon emissions, an increase in temperature due to climate change can also trigger forest fires [27].

Challenges

- A Weak visual interpretation, which cannot distinguish between burned and unburned trees. A proper classification method is needed, which might aid in estimating damaged areas and developing a recovery strategy [30].

- Achieving the required high temporal resolution for wildfire detection and necessitating near-real-time monitoring. The present low temporal resolution makes then the detection of fire not as near real-time as possible and leads to an inevitable lag in the dissemination and depiction of data [31].

- Detection of smoke plumes generated by fire emission: Image processing services do have significant drawbacks such as the lack of prediction of the incident plus some biasing factors like the propagation of smoke is safe and non reached areas. Moreover, smoke is undetectable at night, making the image processing requirements for smoke plume decomposition effort demanding and slow. [32]

- False fire detection due to contextual and thresholding contextual algorithms, high-frequency fluctuations in the fire front, errors in image georeferencing, and environmental components such as clouds and water. [33] [34]

- Hardware issues such as the saturation of the mid-infrared channels of satellite sensors at fairly low temperatures. Furthermore, we have battery life-related issues for UAVs and drones, which is currently an intriguing factor for spatially extensive applications.

- The inclusion of clouds is a prevalent trigger of false alarms in the abovementioned fire detection algorithms. Clouds typically have cool tops, leading to a low brightness temperature. As a corollary, fires below these clouds are not recognized, since temperatures at the top of the sky cannot detect them. That would have a bearing on the analysis, and the algorithm will produce skewed results.

- CubeSats limitations: short-term performance with respect to pointing accuracy and stability. Long term performance including ADCs component robustness and system efficiency after being in orbit for certain period without maintenance [35].

- In the software domain, there is a crucial underestimated data crisis. Data might become incomplete, old, or incorrect, making the results erroneous. Several recent studies have shown that moderate-resolution burned-area products are not able to adequately map the occurrence of small fires due to biased data. [36]

Impact

Investing in such a promising sector as remote sensing will only be but beneficial to societal, financial, ecological, and technological sectors. First, it will help us foresee one of the predominant climate change responses since there is a correlation between temperature rise and fire regimes.
Furthermore, forests present a complex ecosystem with societal and ecological

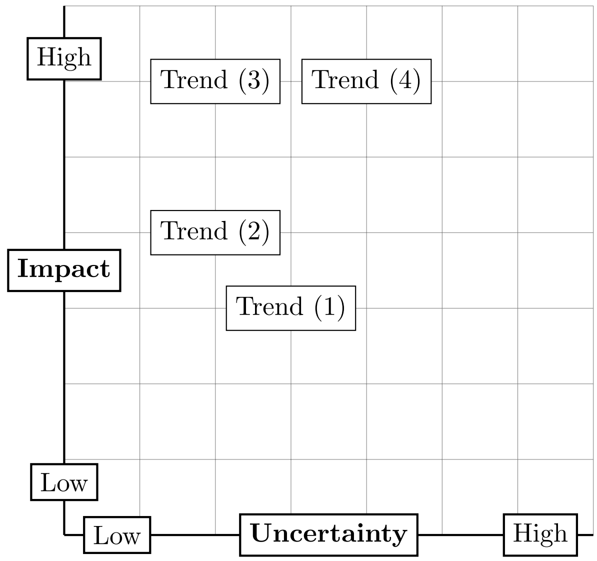

Figure 9.4: Driver matrix

benefits. Mastering their preservation and predicting devastating wildfires can improve sustainable and multi-functional forest management. Relying on RS technologies is not only beneficial in preventing wildfires, but can help determine the extent of forest stand damage, ascertain the ecosystem's potential to regenerate naturally after a fire, aid in the organization of reclamation interventions, determine the dynamics of natural recovery, and decide the future of any eventual restoration intervention. A further impact of fire risk management is to prevent as much as possible the release of carbon into the atmosphere considering forests as huge carbon sponges. For humans, this is also a relief especially for firefighters, who are risking their lives each time without a strategic road map and sacrificing their lives to save others. Huge material casualties can be avoided which strengthens our hope and preserves fire fighting departments. Finally, I would insist on the indirect but predominant impact of remote sensing on a scientific and technological level. Imaging, Big Data, Machine learning, hardware development, and computer vision could be taken to a whole other level since their combination presents an opportunity to work in a reach scientific environment where several issues are tackles from data storage, to optical optimization and cloud computing.

9.3 Conclusion

In this report we wanted to introduce you to Data Science in Forestry and Agriculure. We tried to accomplish that by going through four different trends. First starting with the Sensor-Supported (climate smart) forestry management

9.2.1, then going through embedded system in agriculture 9.2.2, afterwards dicuss big data and IoT in agriclutre 9.2.3 and lastly talk about remote sensing operational systems for wild land fire crisis9.2.4. Every trend is going through *Introduction, Key drivers, Facts, Challenges* and *Impact*. The importance of each trend is dependend on which aspect you consider the most important. It is highly debateable and depending on other factors, but we tried to lay out the trend in a grid divided by uncertainty and impact. The trends in Figure 9.4 are sorted by chronological order, as they appear in the report, so trend 1 correlates with 9.2.1, trend 2 with 9.2.2, trend 3 with 9.2.3 and trend 4 with 9.2.4. As you can see in Figure 9.4 the trend are spread out in impact and uncertainty, but they tend to have high impact with a low uncertainty. The research so far is interesting and promising for the future, but there is much left to be researched, which can have a high impact on our life and enviroment.

References

[1] Chiara Torresan et al. "A new generation of sensors and monitoring tools to support climate-smart forestry practices". In: *Canadian Journal of Forest Research* ja (2021).

[2] Ivan A Janssens et al. "Europe's terrestrial biosphere absorbs 7 to 12% of European anthropogenic CO2 emissions". In: *science* 300.5625 (2003), pp. 1538–1542.

[3] Interinstitutional File. "Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free movement of Such Data (General Data Protection Regulation)". In: *General Data Protection Regulation* ().

[4] Gert-Jan Nabuurs et al. "By 2050 the mitigation effects of EU forests could nearly double through climate smart forestry". In: *Forests* 8.12 (2017), p. 484.

[5] Jakub Lev et al. "Electrical Capacitance Characteristics of Wood Chips at Low Frequency Ranges: A Cheap Tool for Quality Assessment". In: *Sensors* 21.10 (2021), p. 3494.

[6] Amy Forde. *Farmers save on average 8,700EUR through Smart Farming.* `https://www.farmersjournal.ie/farmers-save-on-average-8-700-through-smart-farming-316720`. accessed 31. August 2021.

[7] Neil Hubbard Joint Research Centre (JRC) of the European Commission Pablo J. Zarco-Tejada and Philippe Loudjani (Monitoring Agriculture ResourceS (MARS) Unit H04). "recision Agriculture: An Opportunity for EU-Farmers – Potential Support with the CAP 2014-2020". In: ().

[8] Dmitrii Shadrin et al. "Enabling Precision Agriculture Through Embedded Sensing With Artificial Intelligence". In: *IEEE Transactions on Instrumentation and Measurement* 69.7 (2020), pp. 4103–4113. DOI: 10.1109/TIM.2019.2947125.

[9] Adnan Shaout et al. "An embedded system for agricultural monitoring of remote areas". In: *2015 11th International Computer Engineering Conference (ICENCO)*. 2015, pp. 58–67. DOI: 10.1109/ICENCO.2015.7416326.

[10] Dmitrii Shadrin et al. "Designing Future Precision Agriculture: Detection of Seeds Germination Using Artificial Intelligence on a Low-Power Embedded System". In: *IEEE Sensors Journal* 19.23 (2019), pp. 11573–11582. DOI: 10.1109/JSEN.2019.2935812.

[11] Sophocleous Marios and Julius Georgiou. "Precision agriculture: Challenges in sensors and electronics for real-time soil and plant monitoring". In: *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. 2017, pp. 1–4. DOI: 10.1109/BIOCAS.2017.8325180.

[12] Ishara Sandeepanie. "Big Data Analytics in Agriculture". In: *University of Moratuwa* ja (2020).

[13] Irena Knezevic Kelly Bronson. "Big Data in food and agriculture". In: *Research Article* 20.22 (2020), p. 6442.

[14] Suiqiong Li, Aleksandr Simonian, and Bryan A. Chin. "Sensors for Agriculture and the Food Industry". In: *The Electrochemical Society Interface* 19.4 (2010), pp. 41–46. DOI: 10.1149/2.f05104if. URL: https://doi.org/10.1149/2.f05104if.

[15] Ji-chun Zhao et al. "The study and application of the IOT technology in agriculture". In: *2010 3rd International Conference on Computer Science and Information Technology*. Vol. 2. 2010, pp. 462–465. DOI: 10.1109/ICCSIT.2010.5565120.

[16] Statista Research Department. *Global agricultural IoT market size in 2018 and 2023*. https://www.statista.com/statistics/766793/global-iot-in-agriculture-market-size/. accessed 23. July 2021.

[17] Samudra Vishal Mukherji et al. "Smart Agriculture using Internet of Things and MQTT Protocol". In: *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*. 2019, pp. 14–16. DOI: 10.1109/COMITCon.2019.8862233.

[18] Durai Raj Vincent et al. "Sensors Driven AI-Based Agriculture Recommendation Model for Assessing Land Suitability". In: *Sensors* 19.17 (2019). ISSN: 1424-8220. DOI: 10.3390/s19173667. URL: https://www.mdpi.com/1424-8220/19/17/3667.

[19] Arnab Kumar Saha et al. "IOT-based drone for improvement of crop quality in agricultural field". In: *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. 2018, pp. 612–615. DOI: 10.1109/CCWC.2018.8301662.

[20] G. Arvind et al. "Automated irrigation with advanced seed germination and pest control". In: *2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)*. 2017, pp. 64–67. DOI: 10.1109/TIAR.2017.8273687.

[21] Thomas Truong, Anh Dinh, and Khan Wahid. "An IoT environmental data collection system for fungal detection in crop fields". In: *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*. 2017, pp. 1–4. DOI: 10.1109/CCECE.2017.7946787.

[22] Jash Kotari. "Challenges Faced By IoT in Agricultural Sector". In: *Geeks for Geeks* (2019).

[23] Tanmay Baranwal, Nitika, and Pushpendra Kumar Pateriya. "Development of IoT based smart security and monitoring devices for agriculture". In: *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*. 2016, pp. 597–602. DOI: 10.1109/CONFLUENCE.2016.7508189.

[24] Abdul Mannan et al. "CO 2 emission trends and risk zone mapping of forest fires in subtropical and moist temperate forests of Pakistan". In: *Applied Ecology and Environmental Research* 17 (Mar. 2019), pp. 2983–3002. DOI: 10.15666/aeer/1702_29833002.

[25] Alex M Lechner, Giles M Foody, and Doreen S Boyd. "Applications in remote sensing to forest ecology and management". In: *One Earth* 2.5 (2020), pp. 405–412.

[26] Weitao Zou et al. "A survey of big data analytics for smart forestry". In: *IEEE Access* 7 (2019), pp. 46621–46636.

[27] Sunil Thapa et al. "Forest Fire Detection and Monitoring". In: *Earth Observation Science and Applications for Risk Reduction and Enhanced Resilience in Hindu Kush Himalaya Region*. Springer, 2021, pp. 147–167.

[28] Panagiotis Barmpoutis et al. "A review on early forest fire detection systems using optical remote sensing". In: *Sensors* 20.22 (2020), p. 6442.

[29] Varun Attri, Rajeev Dhiman, and S. Sarvade. "A review on status, implications and recent trends of forest fire management". In: *Archives of Agriculture and Environmental Science* 5 (Dec. 2020), pp. 592–602. DOI: 10.26832/24566632.2020.0504024.

[30] Jung-il Shin et al. "Using UAV multispectral images for classification of forest burn severity—A case study of the 2019 Gangneung forest fire". In: *Forests* 10.11 (2019), p. 1025.

[31] SB Shah et al. "REAL-TIME WILDFIRE DETECTION FROM SPACE–A TRADE-OFF BETWEEN SENSOR QUALITY, PHYSICAL LIMITATIONS AND PAYLOAD SIZE." In: *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* (2019).

[32] Xiaoman Lu et al. "Detection of Fire Smoke Plumes Based on Aerosol Scattering Using VIIRS Data over Global Fire-Prone Regions". In: *Remote Sensing* 13.2 (2021), p. 196.

[33] Chiara Torresan et al. "Forestry applications of UAVs in Europe: A review". In: *International Journal of Remote Sensing* 38.8-10 (2017), pp. 2427–2447.

[34] Jesus San-Miguel-Ayanz and Nicolas Ravail. "Active fire detection for fire emergency management: Potential and limitations for the operational use of remote sensing". In: *Natural Hazards* 35.3 (2005), pp. 361–376.

[35] Jian Guo and Congying Han. "Where Is the Limit: The Analysis of CubeSat ADCS Performance". In: *4S Symposium 2016*. European Space Agency and the Centre National d'Etudes Spatiales. 2016, pp. 1–15.

[36] Niels Andela et al. "The Global Fire Atlas of individual fire size, duration, speed and direction". In: *Earth System Science Data* 11.2 (2019), pp. 529–552.