CLASSIFYING AND INVESTIGATING THE DYNAMIC DRIVERS OF REGIONAL HYDRO-METEOROLOGICAL EXTREME EVENTS IN CLIMATE MODEL ENSEMBLES USING NEURAL NETWORKS

Dissertation zur Erlangung des Doktorgrades an der Fakultät für Geowissenschaften an der Ludwig-Maximilians-Universität München

vorgelegt von

MAGDALENA MITTERMEIER, MÜNCHEN

eingereicht am 10.05.2022, München

Supervisor: Prof. Dr. Ralf Ludwig,
Department of Geography,
Ludwig-Maximilians-Universität München,
Munich, Germany

Second Supervisor: Prof. Dr. Richard Arsenault,
Research Laboratory on Hydrology, Climate and Climate Change,
École de technologie supérieure (ÉTS),
Montréal, Canada

Day of the oral exam: 30.06.2022

"Everything is interconnectedness"
- Alexander von Humboldt

Acknowledgements

Writing these lines, this thesis finally comes to an end. I am looking back at a very exciting, cheerful and also challenging time. During this thesis, I was in the incredible lucky position to be able to follow my scientific interests, dig deeper into the topics that thrive me and grow on the challenges I encountered and the new experiences I made. It was a formative journey that I would never want to miss. This would not have been possible without a number of people supporting me on the way and whom I would like to thank.

First and foremost, I would like to express my deep gratitude to my supervisor Prof. Ralf Ludwig. Thank you for supporting and encouraging me from student-level on, for giving me the opportunity to work on this PhD topic perfectly tailored to my interests, for your professional scientific guidance throughout the way, for our many inspiring conversations on science and beyond, and for constantly supporting and motivating me. In a nutshell, thank you for providing me the best mentoring throughout the entire process of this PhD.

I am also indebted to Prof. Richard Arsenault for his invested time and effort as reviewer of this thesis and part of the jury committee.

Many thanks to my colleagues and companions who made the working environment in our research group so pleasant. Thanks for the refreshing coffee interruptions and always having an open ear. Particular thanks to the ClimEx group: Raul Wood, Andrea Böhnisch and Alexander Sasse. Thank you for the numerous discussions, sharing of python plotting tricks and for the inspiring and productive research environment. Thanks to Raul Wood, Fabian von Trentini and Florian Willkofer, who introduced me in the early phase of my PhD so helpfully to all kind of technical aspects of working with the ClimEx data. Thanks to Vera Erfurth for making my life much easier by handling bureaucracy issues and for always keeping chocolate ready in case a little sugar boost was needed.

I want to thank our colleagues from Ouranos in Montréal for kindly welcoming me

for two research stays in their beautiful city. I remember the visits to the Ouranos consortium as highly inspiring and educational. Particular thanks are to Dominique Paquin, Émilie Bresson, Marco Braun and Martin Leduc who I had the pleasure to collaborate with on scientific publications.

To my parents, family and friends and especially to my husband, Bernhard: thank you for the confidence you always had in me, for the everlasting support and for making it possible for me to embark on this scientific journey and to continue on my path.

Summary

The amount of available data in climate science is increasing rapidly. In addition to satellite and observational data, climate modeling with increasingly complex and high-resolution models is a major contributor to this development. In particular, model ensembles consisting of multiple climate simulations are driving the amount of modeled climate data sharply upward. This includes the research branch of *Single-Model Initial-Condition Large Ensembles (SMILEs)*. These ensembles consist of several model runs of the same climate model driven by the same greenhouse gas concentration scenario, differing only in the slight variation of initial conditions. This makes it possible to represent the natural variations of the climate system due to internal climate variability, an important source of uncertainty in climate modeling. Due to their typically high number of model runs, SMILEs introduce new challenges with respect to efficient data analysis. Techniques from the field of Big Data and machine learning are favorable for this task. Neural networks and deep neural networks from the field of deep learning have recently been increasingly used in climate science to analyze large data sets and to detect non-linear relationships in the data. One field in synoptic climatology for which neural networks have already yielded promising results is the identification and classification of atmospheric patterns involved in the generation of extreme events. Applications of this type fall into the domain of pattern or image recognition and can help to screen large climate model ensembles for the dynamic drivers of extreme events. Examples of such atmospheric drivers include hurricanes, tropical cyclones, and large-scale circulation patterns with high and low pressure systems in the mid-latitudes.

This PhD thesis is dedicated to the classification and analysis of atmospheric drivers of regional hydro-meteorological extreme events such as heavy precipitation, floods, and droughts using machine learning techniques. In the course of four scientific publications, three different atmospheric drivers or groups of drivers are investigated: the Vb-cyclone over Europe, which is often related to heavy precipitation and floods in

the Bavarian region, freezing rain in Montréal (Canada), which is mostly caused by a pressure-driven wind flow along the St. Lawrence River, and finally different circulation patterns over Europe, which are associated with various hydro-meteorological extreme events including drought in Central Europe. In each case, the method optimized for the respective application is presented, an accuracy analysis is performed, and the code is published, which allows the methodology to be applied to other climate data sets. In all cases, meteorological knowledge is integrated into the machine learning based approach. For Vb-cyclones (paper I) and freezing rain (paper II), this is implemented in the form of a two-staged procedure in which neural networks are used to identify circulation patterns before a downstream meteorological analysis is performed. The studies on circulation patterns (papers III and IV) are based on a subjective catalog of historical examples used for training, which incorporates expert meteorological knowledge. In the case of the Vb-cyclones (paper I) and the European circulation patterns (paper IV), a climate change study is conducted to analyze the effects of anthropogenic climate change on the frequency and seasonality of occurrence of these dynamic drivers of extreme events. In each case, a SMILE is used to account for internal climate variability. Using the *Canadian Regional Climate Model version 5 (CRCM5)* and the *Representative Concentration Pathway (RCP)*8.5 scenario, there is no significant change in the number of Vb-cyclones per year between the far future (2070-2099) and the reference period (1980-2009). In seasonality, however, a significant shift is projected, from summer - the season with the highest occurrence of Vb-cyclones in the reference period - to spring. This climate trend is significant and exceeds the noise of internal climate variability. The warm summer months are associated with the highest precipitation sums compared to the other seasons. In the far future, increases in absolute precipitation sums related to Vb-cyclones are projected in all seasons. The European circulation patterns are analyzed in the SMILE of the *Swedish Meteorological and Hydrological Institute (SMHI-LENS)* under the *Shared Socioeconomic Pathway SSP37.0* scenario. Here, small absolute changes in the frequency of occurrence of $\pm$ 5 days per year between the far future (2071-2100)

and a reference period (1991-2020) are shown. For rare circulation patterns, this corresponds in some cases to a high relative change of ± 50 %. Overall, a significant frequency change is shown for 20 of 29 circulation patterns under the consideration of internal climate variability.

Three general recommendations are made for further applications of machine learning in spatial pattern recognition of atmospheric circulation patterns. The first is the integration of meteorological expertise into the study design. This is especially relevant if deeper process understanding is to be generated. Furthermore, it is recommended to pay attention to a balanced test set and to make the test set composition transparent in the course of publication. Especially in climate science and the analysis of extreme events, there are often time series structures and imbalanced class distributions that have to be taken into account. Finally, the third recommendation refers to the scientific work process. While a fully trained neural network allows for efficient, fast data analysis, the work steps to finalize the trained model can be time-consuming and potentially require multiple, manual optimization attempts. Therefore, it is advantageous to train a neural network that allows the identification of several different atmospheric drivers of hydro-meteorological extreme events. As in the case of circulation patterns, such a trained network can be used for several research projects.

Zusammenfassung

Die Menge an verfügbaren Daten in den Klimawissenschaften nimmt rasant zu. Zu dieser Entwicklung trägt, neben Satelliten- und Beobachtungsdaten, vor allem die Klimamodellierung mit immer komplexeren und hochaufgelösteren Modellen bei. Vor allem Modellensembles, die aus mehreren Klimasimulationen bestehen, treiben die Menge an modellierten Klimadaten stark nach oben. Dazu zählt der Forschungszweig, der mit *Single-Model Initial-Condition Large Ensembles (SMILEs)* arbeitet. Diese Ensembles bestehen aus mehreren Modellläufen des gleichen Klimamodells, die mit demselben Treibhausgaskonzentrationsszenario angetrieben werden und sich lediglich in der leichten Variierung der Startbedingungen unterscheiden. Dadurch ist es möglich die natürlichen Schwankungen im Klimasystem aufgrund der internen Klimavariabilität abzubilden, einer wichtigen Unsicherheitsquelle in der Klimamodellierung. Durch ihre typischerweise hohe Anzahl an Modellläufen bringen SMILEs neue Herausfoderungen bezüglich einer effizienten Datenanalyse mit sich. Hierfür bieten sich Techniken aus den Bereichen des *Big Data* und des maschinellen Lernens an. Neuronale Netze und tiefe neuronale Netze aus dem Bereich des *Deep Learnings* werden in jüngerer Zeit verstärkt in den Klimawissenschaften eingesetzt, um große Datensätze auszuwerten und nicht-lineare Zusammenhängen in den Daten aufzuspüren. Ein Bereich der synoptischen Klimatologie, für den neuronale Netze bereits vielversprechende Ergebnisse geliefert haben, ist die Identifikation und Klassifikation von atmosphärischen Mustern, die an der Entstehung von Extremereignissen beteiligt sind. Anwendungen dieser Art fallen in den Bereich der Muster- bzw. Bilderkennung und können dazu beitragen große Klimamodellensembles auf die dynamischen Treibern von Extremereignissen hin zu durchforsten. Beispiele für solche atmosphärischen Treiber sind Wirbelstürme, tropische Zyklone, sowie größräumige Zirkulationsmuster mit Hoch- und Tiefdrucksystemen in den mittleren Breiten.

Diese Doktorarbeit widmet sich der Klassifikation und Analyse der atmosphärischen Treibern von regionalen hydro-meteorologischen Extremereignissen wie Starknieder-

schläge, Hochwasser und Dürre mithilfe von Techniken des maschinellen Lernens. Im Zuge von vier wissenschaftlichen Publikationen werden drei verschiedene atmosphärische Treiber bzw. Gruppen von Treibern untersucht: die Vb-Zugbahn über Europa, die mit häufigen Starkniederschlägen und Hochwassern im bayerischem Raum zusammenhängt, gefrierender Regen in Montréal (Kanada), der meist durch eine druckbedingte Windströmung entlang des St. Lorenz-Stroms verursacht wird und schließlich verschiedene Großwetterlagen in Europa, die mit diversen hydro-meteorologischen Extremereignissen in Verbindung stehen, u. a. mit Dürre in Zentraleuropa. In den Publikationen wird jeweils die, auf die jeweilige Anwendung optimierte, Methode vorgestellt, eine Genauigkeitsanalyse durchgeführt und der Code publiziert, der eine Anwendung der Methodik auf weitere Klimadaten ermöglicht. In allen Fällen wird meteorologisches Wissen in den, auf maschinellem Lernen basierenden, Ansatz integriert. Für Vb-Zugbahnen (paper I) und gefrierenden Regen (paper II) wird das in Form eines zweistufigen Verfahrens umgesetzt, bei der neuronale Netze zur Identifikation der Zirkulationsmuster eingesetzt werden, bevor diese einer nachgeschalteten meteorologischen Analyse unterzogen werden. Die Studien zu Großwetterlagen (paper III und IV) beruhen auf einem subjektiven Katalog an historischen Beispielen, der für das Training verwendet wird und bei dessen Erstellung meteorologisches Expertenwissen eingeflossen ist. Im Fall der Vb-Zugbahnen (paper I) und der europäischen Großwetterlagen (paper IV) wird zudem eine Klimawandelstudie durchgeführt, die Effekte des anthropogen verursachten Klimawandels auf die Häufigkeit und Saisonalität des Auftretens dieser dynamischen Treiber von Extremereignissen analysiert. Dabei wird jeweils ein SMILE verwendet, um die interne Klimavariabilität zu berücksichtigen. Unter Verwendung des *Canadian Regional Climate Model version 5 (CRCM5)* und des *Representative Concentration Pathway RCP8.5*-Szenarios zeigt sich keine signifikante Veränderung der Anzahl von Vb-Zugbahnen pro Jahr zwischen der fernen Zukunft (2070-2099) und der Referenzperiode (1980-2009). In der Saisonalität hingegen wird eine deutliche Verschiebung projiziert, vom Sommer - als der Jahreszeit mit dem höchsten Vorkommen von Vb-Zugbahnen in der Referenzperiode - ins Frühjahr.

Dieser Klimatrend ist signifikant und übertrifft das Rauschen aufgrund der internen Klimavariabilität. Die warmen Sommermonate gehen im Vergleich zu den anderen Jahreszeiten mit den höchsten Niederschlägen einher. In der fernen Zukunft werden Zunahmen der absoluten Niederschläge während Vb-Zugbahnen in allen Jahreszeiten projiziert. Die europäischen Großwetterlagen werden im SMILE des *Swedish Meteorological and Hydrological Institute (SMHI-LENS)* unter dem *Shared Socioeconomic Pathway SSP37.0*-Szenario analysiert. Hier zeigen sich kleine absolute Veränderungen in der Häufigkeit des Auftretens von ± 5 Tagen pro Jahr zwischen der fernen Zukunft (2071-2100) und einer Referenzperiode (1991-2020). Bei seltenen Großwetterlagen entspricht das teilweise einer hohen relativen Änderung von ± 50 %. Insgesamt zeigt sich für 20 von 29 Großwetterlagen eine signifikante Häufigkeitsveränderung unter Berücksichtigung der internen Klimavariabilität.

Für weitere Anwendungen von maschinellem Lernen im Bereich der räumlichen Mustererkennung von atmosphärischen Zirkulationsmustern werden drei grundsätzliche Empfehlungen ausgesprochen. Zum Einen die Integration von meteorologischem Fachwissen in den Studienaufbau. Das ist vor allem relevant, wenn eine Generierung von tieferem Prozessverständnis erzielt werden soll. Darüberhinaus wird empfohlen auf ein balanciertes Testset zu achten und die Zusammensetzung dessen im Zuge der Publikation transparent zu machen. Besonders in den Klimawissenschaften und bei der Analyse von Extremereignissen bestehen häufig Zeitreihenstrukturen und unausgewogene Klassenverteilungen, die zu berücksichtigen sind. Die dritte Empfehlung bezieht sich schließlich auf den wissenschaftlichen Arbeitsprozess. Während ein fertig trainiertes neuronales Netzwerk eine effiziente, schnelle Datenanalyse ermöglicht, können die Arbeitsschritte bis zur Fertigstellung des trainierten Modells zeitaufwändig sein und potentiell mehrere, manuelle Optimierungsversuche erfordern. Von Vorteil ist deshalb das Training eines neuronalen Netzwerkes, das die Identifikation mehrere verschiedener atmosphärischer Treiber von hydro-meteorologischen Extremereignissen ermöglicht. Wie im Falle der Großwetterlagen ist ein solches trainiertes Netzwerk für mehrere Forschungsvorhaben einsetzbar.

Contents

List of Figures

List of Acronyms

AR Assessment Report

CanESM2-LE Canadian Earth System Model version 2 Large Ensemble

CMIP Coupled Model Intercomparison Project

CNN Convolutional Neural Network

CRCM5 Canadian Regional Climate Model version 5

CRCM5-LE Canadian Regional Climate Model version 5 Large Ensemble

ESM Earth System Model

GCM General Circulation Model

HPC High Performance Computing

IPCC Intergovernmental Panel on Climate Change

MPI-GE Max Planck Institute Grand Ensemble

NeurIPS Neural Information Processing Systems

RCM Regional Climate Model

RCP Representative Concentration Pathway

RGB Red-Green-Blue

RNN Recurrent Neural Network

SMHI-LENS Swedish Meteorological and Hydrological Institute - Large Ensemble

SMILE Single-Model Initial-Condition Large Ensemble

SSP Shared Socioeconomic Pathway

1 Introduction

Climate science has become rich of data; data collected by looking back into the past, by observing presently ongoing developments, and by projecting into possible futures. This is illustrated by the fact that the amount of data in climate science is increasing at an unprecedented rate (Schnase et al. 2016). Besides a growing amount of observational, satellite and reanalysis data, especially the tremendous increase in complexity and size of numerical climate simulations has led to what is described as an "explosion" in climate data (Overpeck et al. 2011). For the future, a further increase in volume and complexity of climate data is to be expected (Overpeck et al. 2011; Knüsel et al. 2019). In part, complex earth system models and a high spatial and temporal resolution contribute to this. But also the employment of ensemble techniques largely increases the volume of climate data, as it is the case for the emerging research field of *Single-Model Initial-Condition Large Ensemble (SMILEs)*. SMILEs are an experimental setup in climate modeling with the goal of studying the internal climate variability of the climate system by keeping the external forcing stable while slightly varying the initial conditions. They usually consist of dozens of simulations of a single climate model and thus produce a particularly large data volume. Working with large ensembles introduces new challenges to the climate and impact research community and missing tools for handling such data volumes can lead to a phenomenon called *ensemble fatigue* (Benestad et al. 2017). This term describes the situation when huge volumes of climate data from ensembles are available, but only a small fraction of it is being analyzed due to limited capacities. This leads to missed scientific opportunities as the potential for further insights contained in the data remains unexploited. However, the knowledge gain in climate research largely depends on the quality of the assimilation and analysis of the vast amount of produced data (Schnase et al. 2016).

This calls for the employment of powerful methods for climate data analysis. For this reason, typical methods from Big Data analysis are now being used in climate

science and high hopes are being set in this regard on machine learning techniques (Knüsel et al. 2019; Rolnick et al. 2019; Huntingford et al. 2019). They generally benefit from large training data sets and, once trained, their application on new data takes little computing time. For example, the machine learning technique of neural networks is capable of learning complex non-linear relationships in the data (Liu et al. 2016; Molina et al. 2021). For these reasons, neural networks are considered to have a high potential for solving challenging tasks in climate science including the detection and classification of extreme weather and climate features in spatio-temporal data (Grotjahn et al. 2016; Rolnick et al. 2019; Molina et al. 2021).

Research in the field of synoptic climatology investigates the relation between weather conditions or extreme events and large-scale atmospheric features that are prevailing at this time (Barry 2005), e.g., circulation patterns or cyclones triggering heavy precipitation or atmospheric blocking associated with heat waves and dry periods. Atmospheric, synoptic features are important drivers of extreme events (Sillmann et al. 2017). Here, neural networks can help to identify and classify these synoptic drivers of extreme events in climate data.

Hydro-meteorological extreme events like heavy precipitation, floods, and droughts pose a high risk to the affected society (Farinosi et al. 2020). Understanding how climate change influences the occurrence of their atmospheric drivers is of high societal relevance. Especially on the regional scale, climate and extreme events are - in addition to thermodynamic processes - strongly influenced by atmospheric dynamics (Shepherd 2014; Seneviratne et al. 2021). For example, Europe is under the atmospheric influence of the jet stream with its related (anti-)cyclones (Huguenin et al. 2020). However, the dynamic component of climate change (dealing with the atmospheric circulation) is highly variable (Shepherd 2014; Seneviratne et al. 2021). There is a lack of knowledge about the key characteristics, precursors, and changes of dynamic drivers of extreme events (Chattopadhyay et al. 2020). This brings along a high uncertainty in climate projections on regional extreme events (Horton et al. 2015; Pfahl et al. 2017; Chen et al. 2018). A better understanding of the dynamic

drivers of regional hydro-meteorological extreme events remains an open research challenge. One important source of uncertainty in dynamic aspects of regional climate change is internal climate variability (Shepherd 2014; Pfahl et al. 2017). Therefore, SMILEs can play an important part in separating climate change induced changes in the occurrence of dynamic drivers of extreme events from the noise of internal climate variability. Given the large data volumes that come along with SMILEs, the benefits of neural networks come into play at this point for efficient data analysis. Neural networks have successfully been applied for pattern recognition tasks in climate sciences (Racah et al. 2017; Chattopadhyay et al. 2020). They can help to classify and identify regional extreme events according to their dynamic drivers in SMILEs and thus enable one step towards a better understanding of the dynamic component of climate change effects on regional hydro-meteorological extreme events.

Linking this together, this thesis addresses the challenging research topic of investigating climate change effects on regional hydro-meteorological extreme events by looking at the dynamic component. The thesis focuses on detecting, classifying and investigating synoptic phenomena of the atmospheric circulation that are associated with these extreme events in large climate ensembles by means of the machine learning technique of neural networks.

1.1 Single-Model Initial-Condition Large Ensembles (SMILEs)

An emerging research branch in climate modeling, which tremendously increases the amount of climate data being produced, is the field of *Single-Model Initial-Condition Large Ensembles (SMILEs)*. A SMILE consists of several simulations of a single climate model starting with different initial conditions, but using the same forcing scenario. The simulations (ensemble members) are physically equally likely representations of the climate system. The spread among the ensemble members is interpreted as a representation of internal climate variability. SMILEs are valuable tools to separate the climate signal, which is the mean response to external forcing in the ensemble, from the noise of internal climate variability (Maher et al. 2021). The number of necessary climate simulations largely depends on the research question, but usually SMILEs consist of dozens of realizations. As a consequence, the data amount goes far beyond what is typical for multi-model ensembles and quickly reaches the scale of hundreds of Terabytes in total.

1.1.1 Uncertainty Source of Internal Climate Variability

The great benefit of SMILEs lies in the possibility to assess the internal climate variability of the climate system. This is one of following three major sources of uncertainties in climate modeling (Hawkins & Sutton 2011; Deser et al. 2012):
1. forcing (scenario uncertainty), 2. model response (model uncertainty), 3. internal climate variability.
The first uncertainty source of *forcing* comes from unknown future anthropogenic behavior with regard to greenhouse gas emissions. This leads to uncertainties about the scale of future anthropogenic forcing and is usually addressed by using different scenarios, e.g., the *Representative Concentration Pathways (RCPs)* by the *Intergovernmental Panel on Climate Change (IPCC)*. When the same external forcing is presumed, different climate models may still project different changes resulting from that forcing. This is, for example, due to different numerical formulations in the models

and is described by the second uncertainty source of model response. This uncertainty source is analyzed by using multi-model ensembles with different climate models. Internal climate variability, on the other hand, denotes natural random fluctuation of the climate system. This variability arises from non-linear dynamic processes in the atmosphere and ocean and is intrinsic to the climate system (Hawkins & Sutton 2011; Deser et al. 2012). Internal climate variability is usually highest for projections on shorter time scales for the next decades ahead (Hawkins & Sutton 2011; Lehner et al. 2020). Figure 1 shows the spread of 24 members of a SMILE, the *Canadian Regional Climate Model version 5 Large Ensemble (CRCM5-LE)*, for short-term projected changes of December precipitation and illustrates the partly strong deviations in the regional patterns between single members (Leduc et al. 2019). The ensemble members are equally likely and the deviations are entirely due to natural fluctuations.

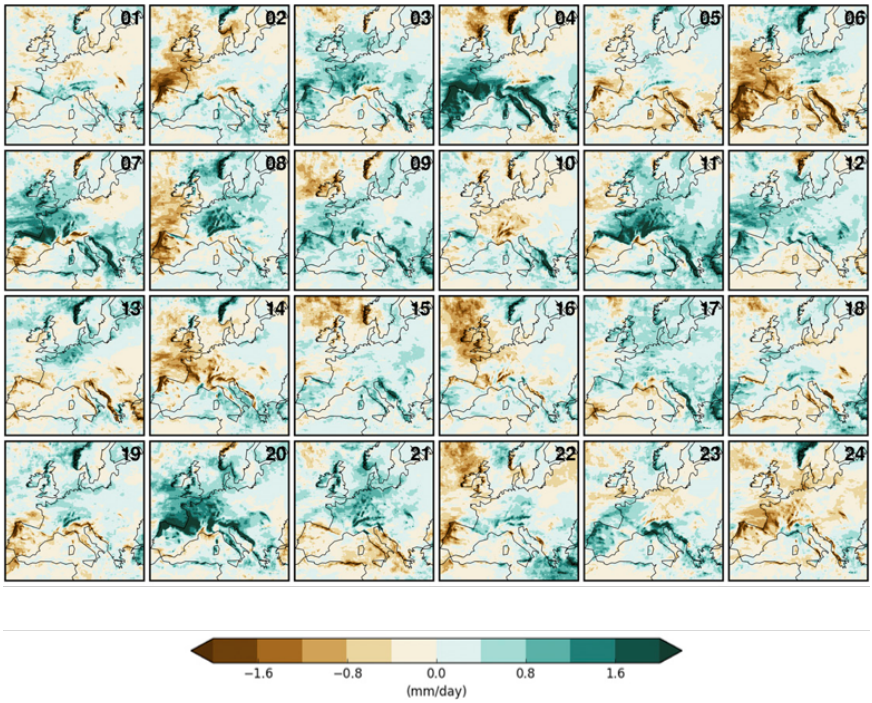

Figure 1: 24 Members of the CRCM5-LE over Europe, which illustrate the spread due to internal climate variability for short-term climate change projections of mean December precipitation (2020–39 vs 2000–19; adapted from Leduc et al. 2019).

1.1.2 Regional SMILEs

SMILEs exist for *General Circulation Models (GCMs)* and *Earth System Models (ESMs)* as well as *Regional Climate Models (RCMs)*. RCM simulations are carried out on a higher spatial resolution and allow a better representation of geographical features such as mountains or land-sea contrasts. They are produced by dynamically downscaling a GCM or ESM. For this, boundary conditions from a global climate model are used over a certain regional domain to drive a RCM, which then evolves its own dynamic and subgrid processes over this limited area in high resolution (Giorgi & Gutowski 2015). Leduc et al. (2019) found that a RCM SMILE better represents local extreme temperature and precipitation than its driving GCM SMILE due to its higher spatial resolution (Maher et al. 2021). Furthermore, internal climate variability increases at smaller spatio-temporal scales (Aalbers et al. 2018; Wood & Ludwig 2020). This underpins the importance of using a RCM SMILE that can resolve finer scales and processes when working on regional phenomena.

Besides the possibility to assess the uncertainty of internal climate variability, another advantage of SMILEs is that they allow a robust sampling of rarely occurring extreme events (Maher et al. 2021). With this, regional SMILEs are an ideal tool for the investigation of regional hydro-meteorological extreme events like heavy precipitation, floods and droughts. However, the large data amount of SMILEs introduces challenges (as mentioned in chapter 1) that call for suitable methods for the handling of scientific Big Data.

1.2 Neural Networks and Deep Learning

The term machine learning in general describes algorithms that enable computers to undergo a learning process without explicitly programming them for the task (Silva & Zhao 2016). Various types of algorithm exist for diverse application options. Machine learning algorithms are categorized in two major groups: supervised and unsupervised learning, depending on whether or not a training data set is used to accomplish the learning task. Supervised learning is the most common form in machine learning (LeCun et al. 2015). In the case of supervised learning, linkages in the data are derived from external training data, while in the case of unsupervised learning the structures intrinsic to the data are unveiled (Silva & Zhao 2016).

Artificial neural networks are a type of machine learning algorithm that were inspired by the biological structure and functionality of the brain of mammals. Key to the brilliant performance of the biological brain is the enormous number of roughly 10^{11} highly interconnected neurons with approximately 10^4 connections per neuron. In a complex chemical process, these neurons pass information to the next neurons. While some of the neuronal structure in the brain is defined at birth, other connections are variable and built through the process of learning as connections are made or broken (Hagen et al. 2014). Inspired by these processes in the biological brain - even though much less complex -, artificial neural networks consist likewise of several layers of nodes (representing the neurons) that are highly interconnected and whose structure determines the function of the network. This structure is called architecture. Each neuron forwards information to the next layer of neurons with differing intensities. This corresponds to neurons in the brain, which fire in case they are activated (McCulloch & Pitts 1943).

Artificial neural networks (in the following named shortly: neural networks) are a powerful machine learning tool for various applications, especially for image-like data in the field of computer vision. Here, neural networks are often used for image or pattern recognition in form of a classification task. One common example is face recognition on photographs. In earth- and climate sciences neural networks for pat-

tern recognition are frequently being applied to spatial data sets, e.g., satellite or climate model data.

The term deep learning describes various different architectures of neural networks having in common that they consist of a multi-layer stack of modules (LeCun et al. 2015). There are several different architectures for deep neural networks. For example, the architecture of a *Convolutional Neural Network (CNN)* consists of the modules of convolutional, pooling and fully-connected layers. On the other hand, a *Recurrent Neural Network (RNN)* is aimed for sequential data like in speech recognition and uses, in its basic configuration, a deep architecture of many fully-connected layers with feed-back loops for memory (LeCun et al. 2015). CNNs will be further introduced in chapter 1.2.1.

1.2.1 Fundamentals of Neural Networks and Deep Learning

The fundamentals of basic feed-forward neural network are explained in the following for the common supervised learning task in computer vision of classifying pictures (here: circulation patterns). Therefore, a training data set is required, which consists of training images X (e.g., photographs or here circulation patterns at a certain time step) in conjunction with the corresponding class labels y (e.g., class 1: circulation type 1, class 2: circulation type 2). The learning procedure consists of three steps: Forward propagation, calculation of the loss function and optimization with back-propagation. These three steps are explained in the following and the full learning procedure is illustrated in Figure 2.

Forward Propagation

During training, each training example (X, y) is shown to the network and the network makes a hypothesis h on the predicted class affiliation. This is being done through a procedure called *forward propagation*. Forward propagation involves two steps: 1. matrix multiplication between the nodes of the layers and weight parameters Θ (see equation 1; Goodfellow et al. 2016), and 2. the application of an activation function,

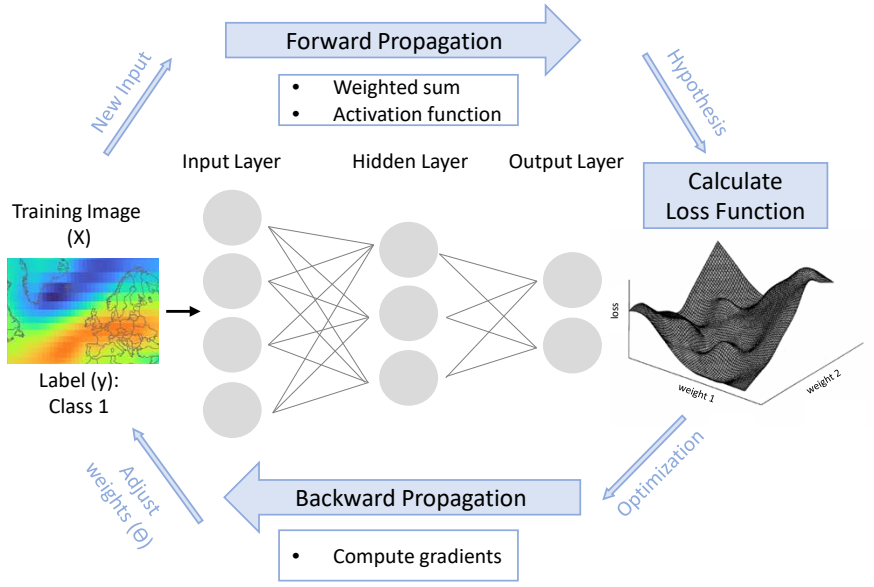

Figure 2: Training cycle of a fully-connected neural network with forward propagation, calculation of loss function and optimization with backpropagation. The training cycle is repeated until a local or global minimum is found. The loss function is derived from Gardner & Dorling (1998).

e.g., the sigmoid function (see equation 2; Hagen et al. 2014). The nodes in the input layer correspond to the pixels of the image. The weights are internal parameters that are randomly initialized at the beginning. In the case of a simple linear model, the weights Θ consist of w and b. A typical activation function for the output layer is the sigmoid function, which gives a binary output into 0 or 1 (Goodfellow et al. 2016).

$$h = X^T w + b \qquad (1) \qquad\qquad f(h) = \frac{1}{1 + e^{-h}} \qquad (2)$$

Loss Function

The class affiliation predicted by the network is then compared with the true class label of the training example by applying a loss function. The loss function depends on the

weights and its output indicates the error the network makes on the training example (see equation 3, with i being a single training example; adjusted based on Goodfellow et al. 2016). Next, the loss function is minimized by means of an optimization algorithm, e.g., gradient descent with backpropagation.

$$J(\Theta) \approx (y_i - h_i)^2 \tag{3}$$

Optimization with Backpropagation

During gradient descent, the partial derivatives of the loss function $J(\Theta)$ with respect to the weights Θ are calculated. The gradients of the partial derivatives are then determined during *backpropagation*. The name backpropagation refers to the backward flow through the network (Rumelhart et al. 1986). Using the gradients, the weights Θ of the network are updated. Then, the next training example is given as input to the network and the cycle of forward propagation, loss function and optimization with backward propagation is repeated. One cycle corresponds to one step of gradient descent towards a local minimum. The cycles are repeated iteratively until a local or global minimum is found. Once a minimum is found, the final weights Θ are determined and the training process of the network is completed. In the case of stochastic gradient descent, one cycle is being done per training example as explained above. Often, mini-batch gradient descent is used for batches sizes of 2^n, where one mini-batch is used for one cycle (Goodfellow et al. 2016). Once the training is completed, the network is ready to be applied to unseen data in order to carry out the classification task. An independent test set, with examples that were not used during the training process, is used to evaluate the accuracy of the trained network and its capability to generalize on new, unseen data (Lones 2021). Several different metrics exist. Besides the overall accuracy, which measures the percentage of correct classifications in the test set, typical measures are the *F1-score* and the *Matthews correlation coefficient* (Matthews 1995), which both consider precision and recall and with this false positive and false negative classifications. For this reason they are more suitable in the case of imbalanced class distributions than overall accuracy.

Convolutional Neural Networks (CNN)

In the 1980s, LeCun (1989) developed a special type of neural networks, called *Convolutional Neural Networks (CNN)*. CNNs use in some of their layers a linear operation called convolution instead of a normal matrix multiplication. A second type of layer that is essential for CNNs is pooling. These two layer types are located in the upper part of the network before one or several fully-connected layers (see Figure 3). CNNs are mainly used for pattern recognition tasks on image-like data (O'Shea & Nash 2015). They are designed for multi-array inputs, e.g., a colored picture with three 2D-arrays containing the pixel values of the red, green and blue (RGB) channels (LeCun et al. 2015). The input to a CNN therefore typically has the size (nH, nW, nC) with nH and nW being the height and width of the image in pixels and nC being the number of channels (for RGB: 3). In the convolutional layer, a number of k filters (kernels) are applied to the input image. These filters run over the image like a moving window and detect lines and edges and extract features in the input image. They also break down the complexity of the pattern. The output of the convolutional layer are k feature maps with a smaller number of pixels (nH, nW) than the input. The number of channels nC stays constant. Next, in the pooling layer the complexity of the feature maps is further reduced by applying filters within which, the maximum values in case of *max pooling* are calculated. Before the feature maps are passed to the subsequent fully connected layers they are flattened to a vector.

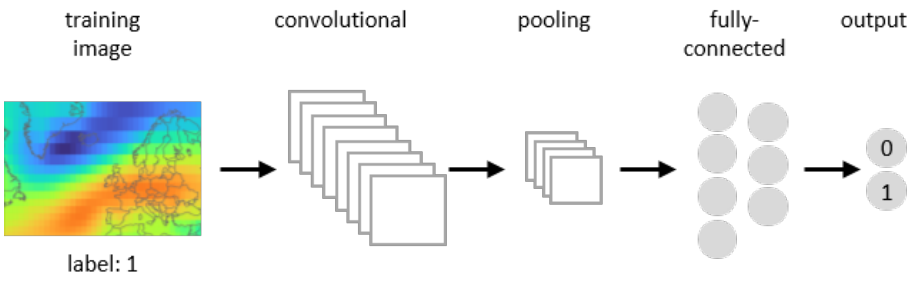

Figure 3: Simple architecture of a CNN with a convolutional, a pooling and fully-connected layers for a binary classification task. The image shows an exemplary circulation type with a high pressure ridge over Europe.

1.2.2 Applications in Climate Sciences: State-of-the-art

Climate science can benefit from the employment of machine learning techniques in numerous ways, especially when combining these techniques with physical domain knowledge (McGovern et al. 2017). One example is the improvement of physically based climate models through machine learning, e.g., by replacing the parameterization of clouds by a deep neural network that is trained on a physical model. The deep neural network is then capable of considering small-scale processes like cloud convection but it is far less computationally expensive than the physically based model and can thus be implemented in a GCM (Gentine et al. 2018). Other examples are the improvement of bias correction through machine learning or the detection of extreme events and the tracking of storms in observational, satellite, forecast or climate model data (Rolnick et al. 2019). Here, machine learning techniques enhance the ability to sift through the vast amount of data (e.g., of large ensembles) in order to extract patterns and insights (McGovern et al. 2017).

Already in the 1990s, there was a wave of machine learning applications in the atmospheric sciences. Gardner & Dorling (1998) report in their review paper about applications for the goal of prediction, function approximation and pattern recognition of satellite images, land cover classifications and atmospheric circulation patterns. Cawley & Dorling (1996) for example reproduced the LAMB catalog, a subjective circulation type classification over the British Islands, using neural networks and achieved a better performance than rule-based classification approaches. Gardner & Dorling (1998) stated that the ability to automate circulation type classifications and to apply them to climate model data using machine learning could improve climate sciences in various ways. Verdecchia et al. (1996) furthermore used neural networks for the detection of winter atmospheric blocking in climate models.

Reasons for the recent breakthrough of machine learning are the tremendous rise of (scientific) data in the last years, the development of more complex machine learning architectures and the more powerful computational environment (Al-Jarrah et al. 2015; LeCun et al. 2015). Kurth et al. (2017), for example, presented the first deep

learning system that was specifically aimed for scientific pattern recognition on Big Data and *High Performance Computing (HPC)* systems, among others the detection of tropical cyclones. Machine learning methods are considered to have a high potential for solving previously challenging tasks in climate sciences including extreme weather classification problems (Rolnick et al. 2019; Huntingford et al. 2019). For such spatio-temporal pattern recognition tasks progress in future research is expected (Grotjahn et al. 2016). The ability of neural networks to capture non-linear information in the data is a key reason for their suitability for such weather and climate detection tasks (Molina et al. 2021). In recent years, few research groups have been focusing on this field (e.g., Racah et al. 2016; Liu et al. 2016; Racah et al. 2017; Chattopadhyay et al. 2018; Lagerquist et al. 2019; Muszynski et al. 2019; Chattopadhyay et al. 2020; Molina et al. 2021). They analyze atmospheric features like synoptic-scale fronts, circulation patterns causing extreme weather, tropical cyclones, extra-tropical cyclones, atmospheric rivers and convective storms. The research goals in analyzing these atmospheric features span from extreme event forecasting, early warning and risk management (e.g., Chattopadhyay et al. 2020; Racah et al. 2017) over climate change studies on future extreme events (e.g., Molina et al. 2021; Muszynski et al. 2019) to create a better understanding about the atmospheric processes leading to an extreme event (e.g., Chattopadhyay et al. 2020). With this, the research topics addressed by employing machine learning techniques for the identification and investigation of atmospheric circulation patterns span a broad range of relevant tasks in studying extreme events.

1.3 Climate Change Impacts on Regional Hydro-Meteorological Extreme Events

Hydro-meteorological extreme events like heavy precipitation, floods and droughts are potential high-impact events that pose a high risk to the affected society (Farinosi et al. 2020). As chapter 11 of the sixth Assessment Report (AR) by the IPCC (Seneviratne et al. 2021) states, changes in extreme events have been observable in recent decades. Under climate change conditions, the frequency and intensity of hydro-meteorological extreme events is projected to increase in many regions on the globe. While there is high certainty about projected changes on a global scale, large uncertainties remain regarding regional trends. This is due to the dynamic drivers of hydro-meteorological extreme events, which are highly complex, spatially very inhomogeneous and remain an uncertain aspect of regional climate change projections (Pfahl et al. 2017; Seneviratne et al. 2021). Changes in regional extremes are furthermore modified by regional forcing and feedbacks like land cover and land use change or aerosol emissions (Seneviratne et al. 2021).

1.3.1 Observed and Projected Changes

Heavy precipitation in terms of the annual maximum daily precipitation (RX1day) has significantly increased since the 1950s in wet but also in dry regions (Dunn et al. 2020; Seneviratne et al. 2021). Floods and droughts are regionally highly variable and the interplay of the driving mechanisms is complex. While there is an observed increase in flood frequency in some regions, it decreases in others. Droughts in the form of precipitation deficits have been observed in hot-spot regions such as Australia, California and the Amazon (Seneviratne et al. 2021). There is *medium confidence* that not only anthropogenic forcing in the form of greenhouse gases but also anthropogenic aerosol emissions, land use changes, and land cover changes influence precipitation and hydro-meteorological extreme events (Seneviratne et al. 2021; *medium confidence* relates to categories by the IPCC).

With progressing climate change extreme precipitation events will very likely get more intense and more frequent over most of the mid-latitude land masses and wet tropics (Seneviratne et al. 2021). While the picture is clear with respect to extreme precipitation, the hydrological response in terms of flooding is much more uncertain (Brunner et al. 2021). For floods, there is *medium confidence* on global trends but *low confidence* in regional projections. Generally, floods are projected to increase over half of the globe. Decreasing trends are found for central and eastern Europe and the Mediterranean (Seneviratne et al. 2021). The study by Brunner et al. (2021) introduces a locally-defined extremeness threshold, which determines the flood response to changes in rainfall extremes for the study area of *Hydrological Bavaria* in southern Germany, which captures the entire catchments of Bavarian rivers and is also partly analyzed in this thesis. The study underlines the variability of the hydrological response depending on the characteristics of the catchment, the season and the event type. The study furthermore confirms the general finding that the more extreme a precipitation event, the greater and more consistent is the response in terms of flood magnitude (Bertola et al. 2020; Brunner et al. 2021). For droughts, there is *high confidence* on a projected increase in drought frequency and intensity in several regions of the globe (Seneviratne et al. 2021). Droughts are classified into several sub-types: meteorological droughts described by a deficit of precipitation, agricultural and ecological droughts considering biological aspects (e.g., soil water deficits), and hydrological droughts associated with reduced water supply. Particularly strong increasing trends in the frequency of meteorological droughts are projected for the regions of the Mediterranean, Southern Africa, Australia, Central-America and parts of South-America (Seneviratne et al. 2021). For Europe, the results by Böhnisch & Mittermeier et al. (2021; *shared first-authorship*) suggest to consider the Mediterranean, the Iberian Peninsula, the Alps and France as hot-spot regions in Europe using the regional SMILE of the CRCM5 under the RCP8.5 scenario.

If humanity manages to comply with the Paris Climate Protection Agreement of limiting global warming to well below 2 °C, the percentage of the global population

affected by hydro-meteorological extreme events would be > 50 % less in Africa, Asia and America and 40 % less in Europe and Oceania compared to a 3 °C scenario (Farinosi et al. 2020).

1.3.2 Thermodynamic and Dynamic Drivers

Projected changes in extreme events caused by anthropogenic global warming underlie a certain variability and uncertainty. Separating thermodynamic and dynamic processes can help to distinguish and better understand the different processes contributing to the generation of an extreme event (Sillmann et al. 2017; Seneviratne et al. 2021). Thermodynamic processes are associated with atmospheric water vapor and are temperature induced, while dynamic processes are related to the atmospheric circulation (Sillmann et al. 2017; Oueslati et al. 2019). Even though the separation helps to generate process understanding, thermodynamic and dynamic processes are highly interconnected and influence each other (Sillmann et al. 2017; Seneviratne et al. 2021).

Synoptic climatology is a research branch that investigates the dynamic component of the climate system and studies the connections between atmospheric circulation patterns and the weather conditions or extreme events associated with these patterns (Barry 2005). On the global scale, there is *high confidence* in the change of extremes - be it in theory, observations or models (Shepherd 2014; IPCC 2013). This is because global changes in extremes are mainly driven by thermodynamic processes and are directly linked to changes in surface temperature. For example, the atmospheric water-holding capacity of air increases with temperature following the Clausius-Clapeyron relation with an increase of 7 % K^{-1}. Analogously, extreme precipitation on the global scale increases with the same rate (Shepherd 2014; Seneviratne et al. 2021). The study by Wood et al. (2021) has quantified the global mean of the average increase in extreme precipitation (99.9th percentile of daily precipitation) in a multi-SMILE ensemble of six SMILES of the *Coupled Model Intercomparison Project version 5 (CMIP5)* with 7.2 % K^{-1} for annual to decadal time scales. Figure 4 shows the

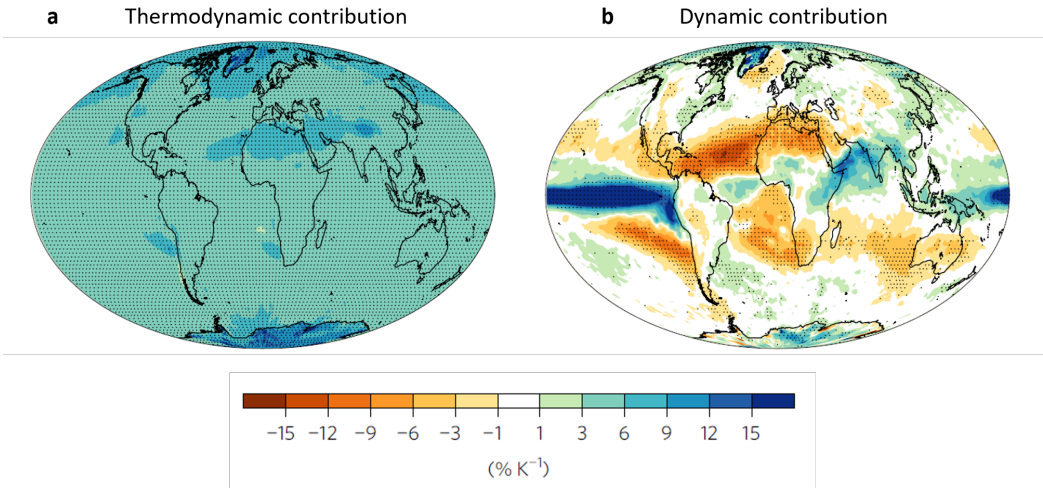

Figure 4: Thermodynamic and dynamic contribution to changes in annual maximum precipitation (RX1day) in 22 CMIP5 models using a physical scaling diagnostic (adapted from Pfahl et al. 2017). **a.** The thermodynamic contribution shows a homogeneous pattern with positive values mainly between $4 - 8$ % K^{-1}. **b.** The dynamic contribution shows regions where the forced response of RX1day cannot be explained by thermodynamics alone, especially in the subtropics, tropics and over oceans.

thermodynamic and dynamic contribution to the change in the RX1day for the entire globe. The thermodynamic processes show a spatially homogeneous picture with an increase in the RX1day that is consistent across climate models. The dynamic contribution, however, is spatially highly inconsistent and partly modifies the regional signals in which it reinforces or counteracts the thermodynamic trend (Pfahl et al. 2017). As a result, the increase in extreme precipitation on the regional scale can show super-Clausius-Clapeyron scaling with increases of $>$ 7 % K^{-1} (Wood & Ludwig 2020). On a regional scale, climate and extreme events are strongly controlled by dynamic processes related to the atmospheric circulation (Shepherd 2014; Pfahl et al. 2017; Seneviratne et al. 2021). For example, in Europe the westerly jet stream of the mid-latitude circulation and the constellation of its cyclones and anticyclones are important drivers of extreme events (Huguenin et al. 2020), e.g., through storm tracks leading to heavy precipitation or atmospheric blocking leading to heat waves or droughts. Such synoptic phenomena (scales of $\geq$ 1000 km; Barry 2005) are relevant

mechanism contributing to the generation of extreme events (Sillmann et al. 2017). It is critical to understand how climate change influences the occurrence of these synoptic drivers of extreme events (Chen et al. 2018). Changes in the atmospheric dynamics, however, are highly complex. They are not robust in observations and remain an uncertain aspect of climate change projections on circulation-related impacts (Shepherd 2014; Pfahl et al. 2017; Seneviratne et al. 2021). Because precipitation is controlled by both thermodynamic and dynamic aspects, regional projections on precipitation related extreme events like floods and droughts have a high uncertainty (Shepherd 2014; Pfahl et al. 2017; Oueslati et al. 2019; Seneviratne et al. 2021).
There is a knowledge gap in climate sciences regarding the understanding of climate change impacts on the atmospheric circulation (Shepherd 2014; Horton et al. 2015; Chen et al. 2018). One of the uncertainty sources for this dynamic component is model error. Especially, insufficient parameterizations of unresolved processes on the sub-grid scale, e.g., clouds and convection, lead to differences in model projections from different climate models and systematic uncertainties in circulation-related projections. Another important source of uncertainty in dynamic aspects of climate change is the internal climate variability (Shepherd 2014; Pfahl et al. 2017). Here, SMILEs can help to statistically distinguish climate change signals from the noise of internal climate variability. The study by Suarez-Gutierrez et al. (2020) for example disentangles the dynamic and local thermodynamic contribution to extreme temperature variability in Europe using the *Max Planck Institute Grand Ensemble (MPI-GE)*. Even though climate change impacts are more robust for climate indicators related to thermodynamics (e.g., surface air temperature) a significant change in risk of circulation-related extremes can still be quantified using SMILEs (Shepherd 2014). Furthermore, there is a research gap on the connection between changes in atmospheric circulation and the resulting consequences on extremes. Since the IPCC AR5 more studies have been focusing on filling this research gap and are addressing the direct connection between changes in extreme-causing circulation patterns and trends in the extremes they are triggering (Chen et al. 2018).

1.4 Research Questions

This thesis addresses the research question of how climate change influences the dynamic drivers of regional hydro-meteorological extreme events by classifying synoptic features (e.g., circulation patterns) that trigger extreme events in large ensembles by employing neural networks.
The scope of this thesis can be categorized into two overarching research questions - one with a focus on the employment of neural networks in climate sciences, the other one with a focus on climate change impacts:

Q1: Can neural networks facilitate the investigation of atmospheric drivers of regional hydro-meteorological extreme events by identifying them in large climate model ensembles?

Q2: How does climate change influence the occurrence of atmospheric drivers of regional hydro-meteorological extreme events? What role does internal climate variability play?

These questions will be addressed on the basis of three synoptic features that are associated with regional hydro-meteorological extreme events:

1. Vb-cyclones over Central Europe, which are a specific type of extra-tropical cyclones whose characteristic pathway makes them an important atmospheric driver of extreme precipitation and floods in the northern Alpine foreland including the study area of *hydrological Bavaria*.
   The respective research questions are the following:

   - *Q1.1: Can neural networks for pattern recognition detect cut-off low related Vb-cyclones in regional climate model data sets? Can this machine learning approach be combined with the established meteorological procedure of cyclone tracking?*

- *Q2.1: How does climate change under the RCP8.5 scenario affect the frequency, seasonality and precipitation intensity of cut-off low related Vb-cyclones? What role does internal climate variability play?*

2. Mixed Precipitation (freezing rain and/or ice pellets) in the study area of Montréal, Canada. Montréal is prone to mixed precipitation due to its location at the St. Lawrence river valley, where under certain circumstances winds get channeled down in the direction towards Montréal and lead to the characteristic air layering that favors the occurrence of mixed precipitation. Freezing rain is one of the costliest high-impact events in Montréal.
   The respective research question is the following:

   - *Q1.2: Can deep learning help to identify the large-scale atmospheric driver of mixed precipitation in Montréal in regional and global climate models? How can this deep learning approach be combined with meteorological domain knowledge?*

3. The circulation types by Hess & Brezowsky over Europe are called *Großwetterlagen* and classify the large-scale atmospheric circulation throughout the year over Europe and parts of the North Atlantic into 29 classes. Six of these 29 circulation types are specifically related to the occurrence of drought and heat. The respective research questions are the following:

   - *Q1.3: Can deep learning help to study important atmospheric drivers of heat and drought in large climate ensembles?*
   - *Q1.4: Can deep learning provide a more accurate classification method of Hess & Brezowsky's circulation types over Europe than the state-of-the-art? What uncertainties are involved in the deep learning approach?*
   - *Q2.2: How does climate change affect the frequency distribution of Hess & Brezowsky's circulation types over Europe in the Swedish Meteorological and Hydrological Institute - Large Ensemble (SMHI-LENS), a SMILE from the CMIP6 generation? What is the role of internal climate variability?*

2 Scientific Publications

This cumulative thesis consists of four publications - three articles in peer-reviewed journals and one workshop paper. The latter was published as preprint in the course of a workshop of the conference on *Neural Information Processing Systems (NeurIPS)*. Two of the peer-reviewed articles have already been published, one has the status of being submitted. The papers are presented in the following. Each is introduced by a concise overview page. This overview page leads from one paper to the next and gives information about the journal, the impact factor, the status of the paper, and the contribution of the authors. The papers are presented in chronological order as this follows the thread of this thesis. All papers deal with dynamic drivers of different regional hydro-meteorological extreme events. They employ large ensembles of climate models either in their methodological design or in the usage of SMILEs to discriminate between climate change effects and internal climate variability. Machine learning techniques are used to deal with the large volumes of climate data involved. This thesis is based on interdisciplinary work in the research field between climatology, meteorology, and computer science. It takes advantage of modern geography's perspective on the research topics covered. From the author's point of view, this primarily consists of the components: interdisciplinarity, interconnectedness, and the regional context of global change. Interdisciplinarity, because it is a skill that geographers are explicitly being trained for (Bracken & Oughton 2009) and that has been used throughout this thesis in form of close collaboration with meteorologists, statisticians, and computer scientists. Interconnectedness due to the complexity of climate change impacts on regional extremes, which are not only driven by thermodynamic but also dynamic factors, which are again subject to a complex nature. An example for this is the influence of Arctic warming on mid-latitude circulation, which is subject to current scientific debates with contrasting hypotheses (Sandu et al. 2016). Finally, regional context of global change contains the *regional embeddedness* from the core definition of geography (Liu et al. 2022). The thesis addresses the impacts of global

climate change in the form of regional extremes, focusing on the regional context while considering the global perspective in formulating the research questions. Figure 5 illustrates the arrangement of publications in the interdisciplinary research field touched upon in this thesis and in modern geography. While the three geographical components are equally important for the entire thesis, publications I and IV deal more with climatology than papers II and III. Paper II has a strong methodological focus and deals with a regional, meteorological phenomenon, which is why two co-authors from meteorology have been involved. Paper I and IV emphasize climate impacts, whereby paper I is closer to meteorology by including a tracking procedure, while paper IV focuses on the uncertainty assessment of the deep learning method.

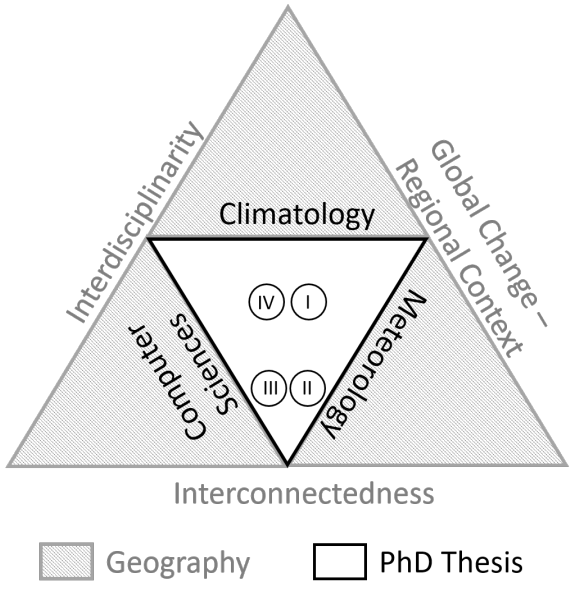

Figure 5: Localization of publications in the interdisciplinary research field of climatology, meteorology and computer sciences and in modern geography. The black-lined triangle representing this PhD thesis is located in the middle of the geography triangle as the three outer components are equally important for this thesis. The papers are located closest to the discipline that was most important for the article.

2.1 Paper I: Detecting Climate Change Effects on Vb Cyclones in a 50-Member Single-Model Ensemble Using Machine Learning

Reference: Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y. & Ludwig, R. (2019): Detecting Climate Change Effects on Vb Cyclones in a 50-Member Single-Model Ensemble Using Machine Learning. *Geophysical Research Letters*. 46 (24), 14653–14661. DOI: 10.1029/2019GL084969.

Transition to paper I: Paper I focuses on the atmospheric feature of a certain type of extra-tropical cyclone, called *Vb-cyclone* (Van Bebber 1891). Vb-cyclones are important dynamic drivers of the hydro-meteorological extreme event of heavy precipitation and floods in the study area of Bavaria. Paper I uses the machine learning technique of neural networks to detect Vb-cyclones in the regional SMILE of the CRCM5-LE from the ClimEx-project (www.climex-project.org). Employing the regional SMILE, the study analyses internal climate variability and forced change on Vb-cyclones. The novelty of the study lies in the accounting for internal variability in the context of Vb-cyclones. The study furthermore combines a machine learning technique with an established meteorological procedure. While the neural network is used to filter the large ensemble with 50 members, cyclone-tracking is applied after the filtering step. In this setting, the data-driven analysis technique of neural networks is integrated in a physically-based modeling and analysis framework.

Author's contribution: MM, MB and RL designed the concept of the study. The study, including all calculations, was carried out by MM, supervised by RL. MH provided the catalog of historic Vb-cyclones. YW supervised the machine learning procedure. MM prepared the manuscript. MB, MH, YW and RL reviewed the manuscript.

Status: published

Journal: Geophysical Research Letters (GRL)

Impact Factor: 4.72

Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL084969

Detecting Climate Change Effects on Vb Cyclones in a 50-Member Single-Model Ensemble Using Machine Learning

M. Mittermeier[1], M. Braun[2], M. Hofstätter[3], Y. Wang[4], and R. Ludwig[1]

[1]Department of Geography, Ludwig-Maximilians-Universität München (LMU), Munich, Germany, [2]Ouranos Consortium, Montréal, Québec, Canada, [3]Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria, [4]Leibniz Supercomputing Centre (LRZ), Garching, Germany

Key Points:
- Machine learning is employed to filter a single-model ensemble for cutoff low related Vb cyclones with high accuracy
- The seasonality of Vb cyclones changes considerably under the RCP8.5 scenario with a peak shift from summer to spring
- The daily precipitation intensity of Vb cyclones significantly increases in all seasons

Supporting Information:
- Supporting Information S1

Correspondence to:
M. Mittermeier,
m.mittermeier@lmu.de

Citation:
Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y., & Ludwig, R. (2019). Detecting climate change effects on Vb cyclones in a 50-member single-model ensemble using machine learning. *Geophysical Research Letters*, *43*. https://doi.org/10.1029/2019GL084969

Received 12 AUG 2019
Accepted 23 NOV 2019
Accepted article online 26 NOV 2019

Abstract Vb cyclones are major drivers of extreme precipitation and floods in the study area of hydrological Bavaria (Germany). When assessing climate change impacts on Vb cyclones, internal variability of the climate system is an important underlying uncertainty. Here, we employ a 50-member single-model initial-condition large ensemble of a regional climate model to study climate variability and forced change on Vb cyclones. An artificial neural network detects cutoff lows over central Europe, which are associated with extreme precipitation Vb cyclones. Thus, machine learning filters the large ensemble prior to cyclone tracking. Our results show a striking change in Vb seasonality with a strong decrease of Vb cyclones in summer (−52%) and a large increase in spring (+73%) under the Representative Concentration Pathway 8.5. This change exceeds the noise of internal variability and leads to a peak shift from summer to spring. Additionally, we show significant increases in the daily precipitation intensity during Vb cyclones in all seasons.

Plain Language Summary Bavaria, a state in the southeast of Germany, has been hit by several devastating floods in recent decades triggered by a storm type called Vb. For future flood risk in Bavaria it is crucial to understand how climate change affects Vb storms. This study uses high-resolution climate simulations over Europe to study changes in the frequency of Vb storms, their seasonal occurrence, and their rainfall intensity under a high greenhouse gas concentration scenario. However, Vb storms are rare events and a single simulation may not provide enough events to distinguish between climate change and random, natural variations. Therefore, we employ a large database of 50 climate simulations with the same settings and greenhouse gas concentration scenario, but slightly different starting conditions, in order to robustly estimate climate change effects on Vb storms. The drawback of using 50 simulations is the high amount of data. Therefore, we apply machine learning for pattern recognition to detect the low-pressure systems related to extreme precipitation Vb storms in the climate simulations. Our results show that climate change considerably affects the seasonal occurrence of Vb storms with a shift from summer to spring. Furthermore, the daily rainfall intensity in Bavaria increases during Vb storms significantly with climate change.

1. Introduction

Climate projections indicate changes in precipitation patterns and an increase in river flood risk in many parts of the world with progressing climate change (Dore, 2005; IPCC, 2012; Willner et al., 2018). Floods count as one of the most devastating natural hazards with a high impact on society (IPCC, 2012; Willner et al., 2018). Midlatitude extreme precipitation and floods are often triggered by certain weather patterns and extratropical cyclones that act as large-scale drivers of intense rainfall (Hofstätter et al., 2016; Messmer et al., 2015). In central Europe, one specific cyclone type, known as *Vb*-cyclone (Van Bebber, 1891), is especially associated with extreme precipitation and a high risk of river flooding (Kundzewicz et al., 2005; Nissen et al., 2013). Vb cyclones develop in the Mediterranean Basin (Gulf of Genoa or northern Adriatic Sea; Muskulus & Jacob, 2005) and propagate to the northeast around the Eastern Alps, leading to orographic precipitation in the northern alpine foreland (Hofstätter et al., 2016; Messmer et al., 2015). One of the regions especially hit by Vb cyclones is the study area of *hydrological Bavaria* in the southeast of Germany (see Figure 1). Although Vb cyclones occur rarely with 2.3 (Messmer et al., 2015) to 4.8 (Hofstätter et al., 2018) events per

Geophysical Research Letters 10.1029/2019GL084969

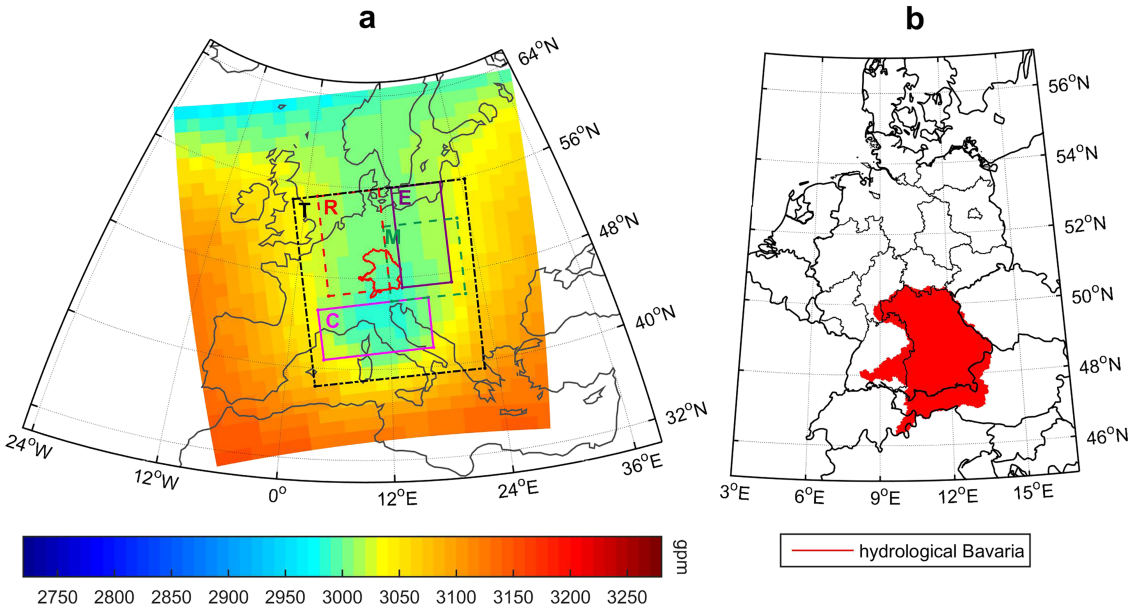

Figure 1. (a) Synoptic pattern at 700 hPa averaged over the 265 Class 1 training examples containing absolute JRA-55 values. A cutoff low over central Europe is visible. The study area of hydrological Bavaria is outlined in red. Additionally, the boxes for the Vb pathway definition are delineated: the tracking zone (T), the cyclogenesis region (C), the end area (E), the mandatory area (M), and the restrictive box (R). (b) State borders (black), federal states of Germany (gray) and position of hydrological Bavaria (red).

year, at least 30% of all summer floods in Bavaria have been related to Vb cyclones (Stahl & Hofstätter, 2018). Vb cyclones that trigger extreme precipitation account for 23% (Messmer et al., 2015) to 41% (Nissen et al., 2013) of all cyclones on a Vb track. Messmer et al. (2015) found the main difference between heavy and weak precipitation Vb cyclones in the corresponding geopotential height field and identified the existence of a cutoff low during heavy precipitation Vb cyclones. Hofstätter et al. (2018) furthermore show that strong Vb cyclones (cyclone intensity >85th percentile) are related to a cutoff low located over central Europe. Climate change is expected to alter the dynamic (e.g., frequency) and thermodynamic factors (moisture transport and related precipitation) of Vb cyclones (Volosciuk et al., 2016). The studies of Volosciuk et al. (2016), Nissen et al. (2013) and Kundzewicz et al. (2005) suggest more intense summer precipitation related to Vb cyclones in a warmer climate. Nissen et al. (2013) further identify a decrease in Vb frequency in summer.

Scientific knowledge on the role of internal variability on changes in Vb cyclone activity is still limited. Internal variability is one of the three major sources of uncertainty in climate change projections, besides model response and external forcing (Deser et al., 2012). It is tied to the intrinsic chaotic character of the climate system due to nonlinear dynamical processes in atmosphere and ocean (Deser et al., 2012). Internal variability affects climate projections, particularly at the regional scale and regarding extremes (Kay et al., 2015; Perkins & Fischer, 2013). This is why the range of internal variability for regional projections regarding rare Vb cyclones is an important open research question. In order to study internal variability, single-model ensembles are employed, which consist of several simulations of the same climate model using slightly different initial conditions (Deser et al., 2012). This paper examines cutoff low-related Vb cyclones using a large single-model ensemble with 50 members of high-resolution climate simulations, the Canadian Regional Climate Model Large Ensemble (CRCM5-LE; Leduc et al., 2019; von Trentini et al., 2019). The drawback of such a large single-model ensemble, however, is the concomitant data amount, which is overall 400 TB for the CRCM5-LE, with several terabytes per variable (Leduc et al., 2018).

This study addresses challenges in identifying Vb cyclones in large climate ensembles by employing machine learning as an efficient data-handling technique. Machine learning proves to be a powerful tool for pattern recognition and seems promising for high-performance analysis of spatiotemporal climate data sets (Grotjahn et al., 2016). Here, a supervised machine learning algorithm for pattern recognition, an artificial

Geophysical Research Letters 10.1029/2019GL084969

neural network (ANN), is employed. Using the ANN, it is possible to scan the CRCM5-LE for the synoptic pattern of a cutoff low over central Europe, which is associated with the initial state of extreme precipitation Vb cyclones (Hofstätter et al., 2018; Messmer et al., 2015). Thus, the large ensemble is specifically filtered for potential cutoff low-related Vb cyclones, before a detailed tracking of their cyclone centers and an analysis of their pathways is conducted. First, cyclones following the Vb pathway are identified, then climate change effects on their occurrence and precipitation intensity over the study area of hydrological Bavaria are studied.

This study addresses the following research question: How does climate change under the Representative Concentration Pathway 8.5 (RCP8.5; Meinshausen et al., 2011) affect the frequency and seasonality of cutoff low-related Vb cyclones and their precipitation intensity while considering the noise of internal variability?

2. Methods

To address this research question on how climate change and internal variability affect cutoff low-related Vb cyclones, a large single-model ensemble with 50 members, the CRCM5-LE, is examined. The CRCM5-LE was created as part of the ClimEx project (Climate Change and hydrological Extremes; Leduc et al., 2019; www.climex-project.org; von Trentini et al., 2019). The CRCM5-LE was generated by dynamically downscaling the 50-member initial-condition ensemble of the Canadian Earth System Model version 2 (CanESM2-LE; Fyfe et al., 2017) using the Canadian Regional Climate Model version 5 (CRCM5; Martynov et al., 2013; Separovic et al., 2013). The dynamical downscaling was performed for two domains (Europe and northeastern North America) to a high spatial resolution of 0.11°. The time series cover the period from 1950 to 2099. Up to the year 2005 the model is forced with historic greenhouse gas and aerosol emissions, while from 2006 on the RCP 8.5 forcing scenario is used (Leduc et al., 2019). The variables examined here are geopotential height at 700 hPa (z700; 3-hourly) and precipitation (pr; hourly; for hydrological Bavaria only). One CRCM5 run was driven by the boundary conditions of the ERA-Interim (Dee et al., 2011) reanalysis (ERA-Interim-CRCM5) and covers the period of 1979 to 2013. It is used to validate the model's capability of reproducing historic Vb cyclones.

The training set for the machine learning algorithm is based on reanalysis data of historic cutoff lows over central Europe and their related Vb cyclones. The Japanese 55-year Reanalysis (JRA-55; Harada et al., 2016; Kobayashi et al., 2015) is used. Dates of historic Vb cyclones are derived from a JRA-55 based catalog of historic cyclone tracks over central Europe from Hofstätter et al. (2018). The catalog identifies 296 Vb cyclones in the period from 1959 to 2015. In the catalog 23% of Vb cyclones occur with associated cutoff lows at 700 hPa, but these account for 75% of extreme precipitation events over hydrological Bavaria associated with Vb cyclones (minimum distance to other cyclones: 500 km). The 69 Vb cyclones with a distinct cutoff low, which consist in total of 265 time steps (training examples), are manually extracted. The JRA-55 reanalysis has a temporal resolution of 6 hr and a spatial resolution of 1.25° on a uniform latitude-longitude grid. The variable used to identify Vb cyclones is z700.

Cutoff low related Vb cyclones in the CRCM5-LE are identified as follows:

1. training an ANN on the detection of historic synoptic patterns of cutoff lows over central Europe with JRA-55 data,
2. applying the trained ANN on the entire ensemble to extract the beginning stage of potential Vb cyclones from the CRCM5-LE,
3. tracking the cyclone centers of potential Vb cyclones, and
4. testing the tracked pathways for fulfilling the definition of Vb.

The following methodology is graphically illustrated in the supporting information (see Figure S2).

First, the training set for the ANN is built based on JRA-55 and the catalog by Hofstätter et al. (2018). The ANN is trained to separate two classes: Class 1 showing a cutoff low over central Europe (positives), and Class 0 showing no cutoff low over central Europe (negatives). For Class 1, the 265 time steps are used that contain a historic cutoff low over central Europe in combination with a Vb cyclone. Figure 1 shows the synoptic pattern at z700 averaged over all positive training examples in JRA-55 reanalysis data. For Class 0 (negatives) 62,955 counterexamples without Vb cyclones are employed. The training data set suffers of a skewed class distribution as historic situations with Vb related cutoff lows over central Europe (positives) are much scarcer than counterexamples (negatives). To account for inaccuracies resulting from such

Geophysical Research Letters 10.1029/2019GL084969

an imbalanced data set, error-weighting, respectively, cost-sensitive learning (Zhou et al., 2018) is applied during training with a weighting factor of 0.025 on Class 0. Preprocessing of JRA-55 data consists of extracting the CRCM5-LE-domain and interpolating the data to the rotated latitude-longitude CRCM5-LE grid (see Figure 1). The original, fine RCM resolution of 0.11° is not necessary for the detection and tracking of large-scale synoptic patterns, thereby data is aggregated to 1.1°.

Second, a two-layered ANN for pattern recognition is employed (input layer: 784 nodes, hidden layer: 25 nodes, output layer: 2 nodes) using MATLAB's Neural Network Toolbox (nntool). The settings of the ANN are the hyperbolic tangent sigmoid function as activation function in the hidden layer (LeCun et al., 2015) and the softmax function in the output layer (Zhou et al., 2018). Cross-entropy (Kline & Berardi, 2005) serves as cost function and is minimized by scaled conjugate gradient backpropagation (Møller et al., 1993). In order to evaluate the performance of the training setup a k fold stratified cross validation is conducted with a number of 10 subsamples ($k = 10$) and a setup division of 90% for training and 5% each for validation and independent test set. As performance indices a confusion matrix with precision and recall is calculated (see Table S1 and Text S1).

Third, the detected events of possible Vb cyclones undergo a tracking procedure. The cyclone centers are tracked based on Murray & Simmonds (1991) without consideration of splitting and merging of tracks. The position of a cyclone center is defined as local minimum in z700 and is tracked during the consecutive time steps of the cyclone's lifetime. For each time step the current cyclone position is identified by first calculating the absolute minimum within a defined distance and, second, comparing this value to its eight neighbor pixels. The distance measure is a constant value of 3 pixels. The tracking ends as soon as the cyclone leaves the tracking area (T) in Figure 1. With this approach, open cyclones that have no clear local pressure minimum and are rather identifiable by local vorticity maxima (Hofstätter et al., 2018) are not detectable. As tracking results also depend on the chosen method (Neu et al., 2013), related uncertainties cannot be resolved in this study.

As the fourth step, the tracked pathways are tested for a set of criteria, which define the characteristics of Vb pathways using four boxes (see Figure 1 Hofstätter et al., 2016; Messmer et al., 2015). Cyclones are classified as Vb if all of the following statements apply: (1) the cyclone track is within the cyclogenesis region (C) for at least one time step, (2) at any later time step the track appears in the end area (E), (3) the cyclone appears within the mandatory area (M) between C and E, (4) the track moves from west to east between C and E, (5) the cyclone does not appear in the restrictive zone (R) any time before it has moved to E and 6. the track lasts for at least 24 hr.

To analyze climate change effects the frequency of Vb cyclones is compared between three 30-year periods: the reference period 1961–1990 (past), 2021–2050 (near future), and 2070–2099 (far future). Significance is tested with a ks test, which is applied to the distribution of the pooled members using a 5% significance level. Additionally, the signal-to-noise ratio (SNR) is calculated according to Aalbers et al. (2018). It describes the ratio between climate change signal and internal variability by dividing the averaged difference between far future and reference (past) by the standard deviation of the CRCM5-LE. $|SNR| > 1$ indicates that the signal has emerged from internal variability and confidence on the sign of change is very high. However, even given a SNR of 1, single members may yield largely deviating changes. At the same time, for an ensemble as large as the one used here, the forced climate response can be robustly estimated even for a SNR smaller than 1.

In order to assess the capability of the CRCM5 to capture the characteristics of Vb cyclones, the CRCM5-LE is compared to the ERA-Interim driven CRCM5 run and ERA-5 reanalysis. ERA-5, the fifth generation of ECMWF atmospheric reanalysis products, is chosen in place of ERA-Interim, because it is available in 3-hourly temporal resolution. The lower temporal resolution of ERA-Interim (6 hr) leads to a noticeably lower accuracy during tracking. The fact that ERA-Interim-CRCM5 is not directly compared to its driving data might, however, be a potential source of inaccuracies in the evaluation.

Additionally, the maximum daily precipitation falling over hydrological Bavaria during a Vb event is analyzed. The maximum daily precipitation values of all Vb events in the ensemble are compared to the 95th percentile of all daily precipitation values (averaged over hydrological Bavaria) in the entire ensemble and the reference period 1980 to 2009 in order to identify the quantity and change of extreme events. The drizzle effect is eliminated by a threshold of <1 mm/day (Kjellström et al., 2010). The value of the 95th percentile is

Geophysical Research Letters 10.1029/2019GL084969

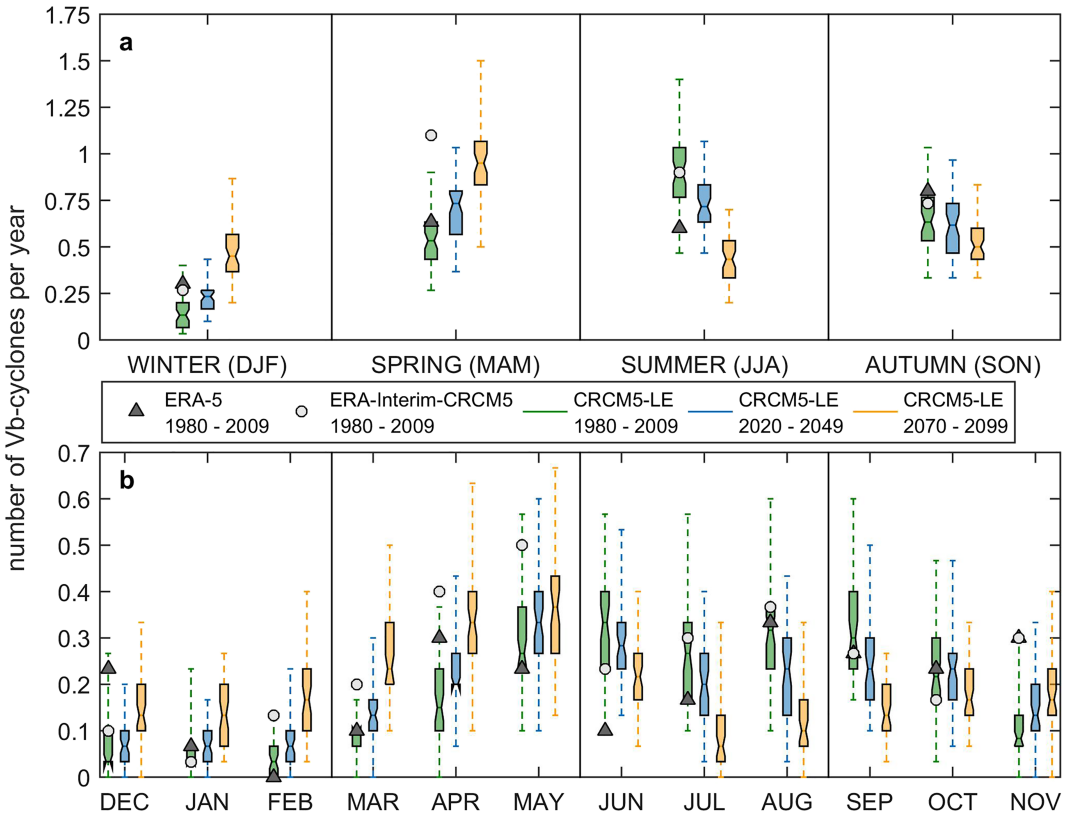

Figure 2. Number of identified cutoff low-related Vb cyclones per year and season (a) respectively per year and month (b) averaged over 30-year periods in the past (1980–2009; green), near future (2020–2049; blue), and far future (2070–2099; orange). The boxplots show the spread between the 50 members of the CRCM5-LE. The line indicates the median, the box stretches from the 25th to the 75th percentiles, and the whiskers extend to the minimum and maximum values. The dark gray triangle (light gray circle) depicts the identified Vb cyclones in ERA-5 reanalysis (ERA-Interim-CRCM5 simulation).

12.7 mm daily precipitation over hydrological Bavaria. According to the seasonality, the precipitation results of the CRCM5-LE are compared to the time series of ERA-Interim-CRCM5 and ERA-5 reanalysis.

3. Results

Vb cyclones connected to a cutoff low over central Europe are identified in the CRCM5-LE with high accuracy indicated by the recall of 94.6% (test set). Table S2 shows the averaged confusion matrices over all cross-validation iterations (see Table S3 for the network used for inference). Due to the skewed class distribution between Class 0 and Class 1, the overall accuracy of 99.2% (test set, Figure S2) is not a reliable indicator for the performance of the ANN in respect of the minority class. The false positives lead to a low precision of 33.2%, but they are minimized in the subsequent step of tracking during which all possible Vb cyclones undergo further tests.

The analysis of cutoff low-related Vb cyclones in the CRCM5-LE under the RCP8.5 scenario indicates a slight, but nonsignificant increase in the absolute number of Vb cyclones per year using the ks test and a 5% significance level (see Figure S3). The mean values in events per year are 2.26 (past), 2.27 (near future), and 2.37 (far future). The spread between the 50 members, which represents internal variability, is considerable. The value of SNR between past and far future is 0.32, which indicates that internal variability is larger than the climate change signal.

Figure 2a shows the number of Vb cyclones per season and year (respectively, Figure 2b per month and year) for three periods of the 50 members of the CRCM5-LE (boxplots) and for the reference period of the single time series of ERA-Interim-CRCM5 and ERA-5 reanalysis (circles and triangles). For the reference

Geophysical Research Letters 10.1029/2019GL084969

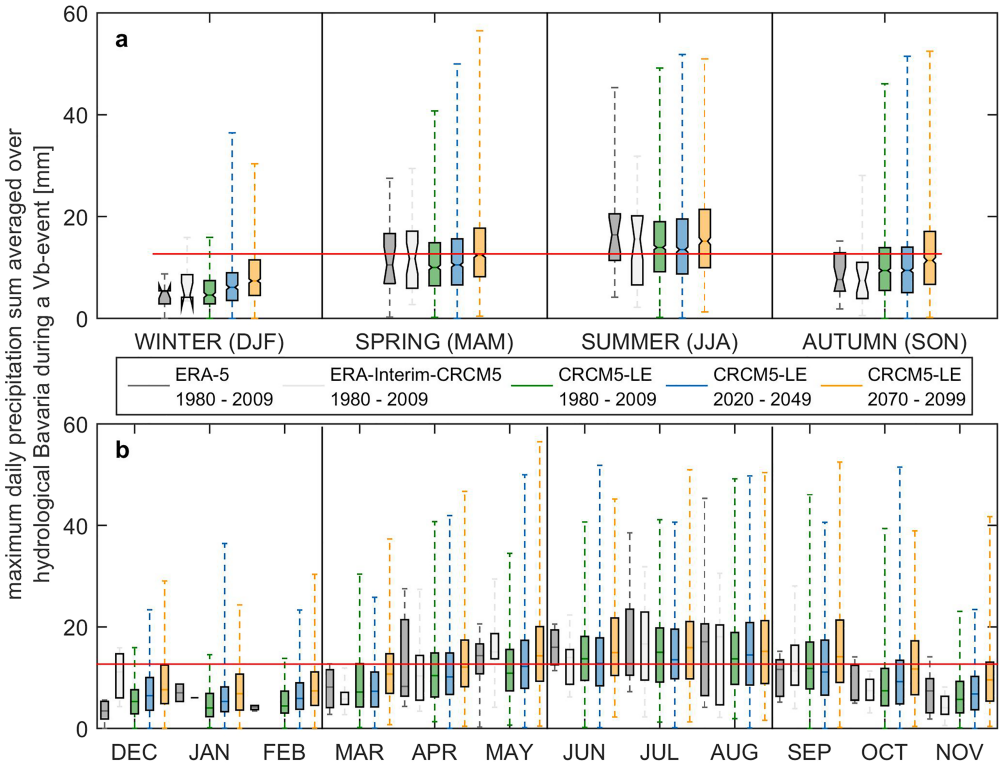

Figure 3. Maximum daily precipitation sum (mm) averaged over hydrological Bavaria during Vb events.The boxplots illustrate the spread of Vb events. For the reference period (1980–2009) ERA-Interim-CRCM5, ERA-5, and the CRCM5-LE are illustrated. The near (2020–2049) and far future (2070–2099) are only covered by the CRCM5-LE. For ERA-Interim-CRCM5 and ERA-5 the underlying database is only a single time series, whereas the CRCM5-LE provides 50 time series.

period the comparison of ERA-Interim-CRCM5 with ERA-5 shows that the frequency of Vb cyclones in the CRCM5 model run, which is driven by ERA-Interim boundary conditions, differs from the frequency of Vb cyclones in the ERA-5 reanalysis product for on-average 0.06 events per month and year. In most months, the frequency of Vb cyclones in the ERA-Interim-CRCM5 time series is higher than in the ERA-5 reanalysis. Here, the CRCM5 model might tend to overestimate the frequency of Vb cyclones. Distinct differences occur especially in spring (May). In two cases in spring, the number of Vb cyclones in the ERA-Interim-CRCM5 run furthermore lies outside of the distribution of the CRCM5-LE (March and April). In most cases, however, the CRCM5-LE covers the time series of both ERA-Interim-CRCM5 and ERA-5.

Looking at the climate change signal between the three 30-year periods of the CRCM5-LE a remarkable impact on the seasonal distribution of Vb cyclones is visible. Whereas in the past, the peak of Vb frequency used to be in summer (June), it transitions to spring (May) with progressing climate change. In the far future, the number of Vb cyclones in summer decreases strongly by −51.8% (mean: −0.47 events/year; changes are calculated between 2070–2099 and the reference period 1980–2009). In the past, one Vb cyclone occurred almost every summer, but in the far future a summer Vb cyclone appears only every 830 days. At the same time the number of Vb cyclones increases considerably in spring by +73.4% (mean: +0.40) to about one Vb cyclone every spring. The changes in spring and summer are significant on a 5% significance level. The SNR values for spring (1.25) and summer (2.86) are clearly larger than 1, consequently the climate change signal exceeds internal variability and confidence in the sign of change is high. The strongest increase happens in the months March and April, the strongest decrease in July and August. Furthermore, Vb frequency significantly increases in winter by +220.7% (mean: +0.32). The change is consistent throughout all winter months. In the far future, the number of winter Vb cyclones even exceeds the frequency in summer. The SNR is 0.82 and therefore noise is larger than the signal. For autumn, the signal is mixed, with a decrease in

Geophysical Research Letters 10.1029/2019GL084969

September and October and an increase in November. Still, there is a slight but significant decrease in Vb frequency in autumn by −20.6% (mean: −0.14). The SNR has a value of 0.67 and thus, as in winter, internal variability is larger than the climate change signal.

Figure 3 shows the seasonality of the maximum daily precipitation averaged over hydrological Bavaria during Vb events. When comparing ERA-Interim-CRCM5 and ERA-5, some differences are clearly visible. For one thing, ERA-Interim-CRCM5 produces lower maximum daily precipitation sums in summer than ERA-5. This is true for both the median values (influence from June) and extreme values (influence from August) and might indicate that the CRCM5 underestimates the maximum precipitation related to Vb cyclones in summer. In autumn (September), however, the ERA-Interim-CRCM5 run contains more extreme values than ERA-5, which implies that the CRCM5 might slightly overestimate the maximum Vb related precipitation in autumn. However, when comparing the single time series of ERA-Interim-CRCM5 and ERA-5 to the reference period of the CRCM5-LE, both distributions are well covered by the large data base of Vb events resulting from the 50 members of the ensemble. The threshold considered for extreme precipitation over hydrological Bavaria (solid red line) is the 95th percentile of all days in the CRCM5-LE reference period. In the past as well as in the future, Vb cyclones mainly exceed this threshold in the warmer months of the summer half year (April to September). In summer the median of all data sets lies above the threshold. For the CRCM5-LE the peak of median values lies in July for both the reference period (15.0 mm) and the far future (15.9 mm). In combination with the results from Figure 2a, this means that the absolute number of Vb cyclones in the season of the highest daily precipitation values (summer) decreases with progressing climate change. At the same time climate change leads to a rise in the maximum daily precipitation sum over hydrological Bavaria in all seasons. Consequently, the percentage of Vb cyclones exceeding the threshold increases in all seasons. The strongest increase in median values between reference and far future occurs in winter with +2.8 mm (spring: +2.4 mm, summer: +1.2 mm, autumn: +2.0 mm). The changes in all seasons are statistically significant on a 5% significance level. It is worth noting that the spread between the Vb events is considerable. Especially on a monthly basis the single time series of ERA-Interim-CRCM5 and ERA-5 do not, in contrast to the CRCM5-LE, provide a statistically reliable distribution of Vb cyclones (e.g. January and February).

4. Discussion and Conclusion

The CRCM5-LE shows that climate change under the RCP8.5 scenario affects the seasonal distribution and rainfall intensity of cutoff low-related Vb cyclones. The absolute frequency of Vb cyclones does not change significantly, but summer Vb cyclones decrease strongly. With this, our results reinforce the findings by Nissen et al. (2013), although we use a large single-model ensemble and examine only one GCM-RCM combination and one scenario. Climate change projections using the CRCM5-LE do not account for uncertainties regarding model or scenario choice, but for the first time allow an in-depth analysis of internal variability regarding Vb cyclones. This shows that the climate change signal of a decrease in summer Vb cyclones exceeds the noise of internal variability. Furthermore, we detected a robust increase in Vb frequency in spring with a SNR larger than 1. In the light of a possible tendency of the CRCM5-LE to overestimate the number of Vb cyclones in certain months, the absolute numbers of Vb events might differ in other model ensembles. The climate change signal between the three periods of the CRCM5-LE, however, is unaffected by a potential model bias. Regarding precipitation intensity related to Vb cyclones, our study shows that climate change leads in all seasons to an increase in the maximum daily precipitation sums over hydrological Bavaria during Vb events. Vb cyclones occur less often in the warm season (summer), which is associated with the highest rainfall intensities. The 50 members of the CRCM5-LE beneficially support the analysis of precipitation related to Vb cyclones, because they provide a larger and thus more reliable database for the analysis of extreme events.

The decrease in summer Vb activity might be due to a shift in the cyclone pathway to the east as suggested by Nissen et al. (2013). Another explanation is a change in Rossby-waves and jet streams due to the Arctic Amplification, which is discussed in the general context of decreasing summer storm activity in the Northern Hemisphere (Mann et al., 2017). The increase in the percentage of Vb cyclones leading to extreme precipitation could be explained by an increased saturation vapor pressure of warmer air following the Clausius-Clapeyron rate and by increased sea surface temperatures of the Mediterranean Sea leading to an enhanced moisture transport by cyclones emerging there (Volosciuk et al., 2016).

Geophysical Research Letters 10.1029/2019GL084969

With a recall of 94.6% Vb cyclones related to a cutoff low over central Europe are successfully detected in the CRCM5-LE using machine learning. Our approach is specifically designed for Vb cyclones that develop during the synoptic situation of a cutoff low over central Europe, because such a cutoff low is an indicator for extreme precipitation Vb cyclones (Hofstätter et al., 2018) and occurred during historic Vb related floods in hydrological Bavaria. With 2.3 Vb cyclones per year identified in the reference period in the CRCM5-LE, the number agrees with the number of Vb cyclones identified by Messmer et al. (2015), though methods and data differ. The number is about half of the Vb cyclones identified by Hofstätter et al. (2018) in JRA-55 data. Hofstätter et al. (2018) also considered open Vb cyclones, which make up to about 50% of all identified Vb cyclones. Limitations of the tracking procedure lie in the incapability of distinguishing splitting tracks or simultaneous cyclones. The choice of 700 hPa as tracking variable has the disadvantage that not all geopotential height minima on a Vb track necessarily extend to the surface or the other way round (Nissen et al., 2013). In contrast, the main advantage of 700 hPa lies in less disturbances due to orography, which leads to less ambiguous cases during tracking (Hofstätter et al., 2018).

Our study provides the novelty of accounting for internal variability in the context of analyzing cutoff low related Vb cyclones with a large single-model ensemble of 50 members. Furthermore, Vb cyclones are simulated by a RCM with 0.11° resolution. This ensures a better representation of orography, a finer delineation of atmospheric processes and of the land-sea contrast (Akhtar et al., 2019; Lucas-Picher et al., 2017), which is important in the development of Vb cyclones. By employing machine learning for cyclone identification an efficient data analysis strategy is ensured in addressing the Big Data scale of the CRCM5-LE.

Acknowledgments
This work was conducted within the ClimEx project (Leduc et al., 2019; www.climex-project.org). Funding from the Bavarian State Ministry for the Environment and Consumer Protection is gratefully acknowledged. We thank BayFOR for funding M. M.'s research stay at Ouranos. The CRCM5 was developed by the ESCER center of UQAM (www.escer.uqam.ca) in collaboration with Environment and Climate Change Canada. We acknowledge CCCma for executing and making available the CanESM2-LE and CanSISE for proposing the simulations. Computations with the CRCM5 for the ClimEx project were made on the LRZ's SuperMUC of BAdW and funded via GCS by BMBF and StMWFK. The cyclone type catalog was derived from M. H. (Hofstätter et al., 2018). JRA-55 reanalysis data were retrieved from JMA (https://rda.ucar.edu). ERA-5 reanalysis data were accessed from Copernicus CDS (https://cds.climate.copernicus.eu/; 2019). Code written in MATLAB R2015a and the trained network are available upon request via zenodo (https://zenodo.org/record/1495000 and https://doi.org/10.5281/zenodo.1480467).

References

Aalbers, E. E., Lenderink, G., van Meijgaard, E., & van den Hurk, B. J. J. M. (2018). Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability? *Climate Dynamics*, *50*(11-12), 4745–4766. https://doi.org/10.1007/s00382-017-3901-9

Akhtar, N., Krug, A., Brauch, J., Arsouze, T., Dieterich, C., & Ahrens, B. (2019). European marginal seas in a regional atmosphere-ocean coupled model and their impact on Vb-cyclones and associated precipitation. *Climate Dynamics*, *53*(9-10), 5967–5984. https://doi.org/10.1007/s00382-019-04906-x

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, *38*(3-4), 527–546. https://doi.org/10.1007/s00382-010-0977-x

Dore, M. H. I. (2005). Climate change and changes in global precipitation patterns: what do we know? *Environment International*, *31*(8), 1167–1181. https://doi.org/10.1016/j.envint.2005.03.004

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., et al. (2017). Large near-term projected snowpack loss over the western United States. *Nature Communications*, *8*, 14996. https://doi.org/10.1038/ncomms14996

Grotjahn, R., Black, R., Leung, R., Wehner, M. F., Barlow, M., Bosilovich, M., et al. (2016). North American extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, *46*(3-4), 1151–1184. https://doi.org/10.1007/s00382-015-2638-6

Harada, Y., Kamahori, H., Kobayashi, C., Endo, H., Kobayashi, S., Ota, Y., et al. (2016). The JRA-55 reanalysis: Representation of atmospheric circulation and climate variability. *Journal of the Meteorological Society of Japan Ser. II*, *94*(3), 269–302. https://doi.org/10.2151/jmsj.2016-015

Hofstätter, M., Chimani, B., Lexer, A., & Blöschl, G. (2016). A new classification scheme of European cyclone tracks with relevance to precipitation. *Water Resources Research*, *52*(9), 7086–7104. https://doi.org/10.1002/2016WR019146

Hofstätter, M., Lexer, A., Homann, M., & Blöschl, G. (2018). Large-scale heavy precipitation over central Europe and the role of atmospheric cyclone track types. *International Journal of Climatology*, *38*, e497–e517. https://doi.org/10.1002/joc.5386

IPCC (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. In C. B. Field et al. (Eds.), *A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (582 pp.). Cambridge, UK, and New York, NY, USA: Cambridge University Press.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. https://doi.org/10.1175/BAMS-D-13-00255.1

Kjellström, E., Boberg, F., Castro, M., Christensen, J. H., Nikulin, G., & Sánchez, E. (2010). Daily and monthly temperature and precipitation statistics as performance indicators for regional climate models. *Climate Research*, *44*(2-3), 135–150. https://doi.org/10.3354/cr00932

Kline, D. M., & Berardi, V. L. (2005). Revisiting squared-error and cross-entropy functions for training neural network classifiers. *Neural Computing and Applications*, *14*(4), 310–318. https://doi.org/10.1007/s00521-005-0467-y

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan Ser. II*, *93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Kundzewicz, Z. W., Ulbrich, U., Brücher, T., Graczyk, D., Krüger, A., Leckebusch, G. C., et al. (2005). Summer floods in central Europe - climate change track? *Natural Hazards*, *36*(1-2), 165–189. https://doi.org/10.1007/s11069-004-4547-6

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Leduc, M., Frigon, A., Brietzke, G., Ludwig, R., Weismüller, J., & Giguère, M. (2018). The ClimEx project: Digging into natural climate variability and extreme events. *Inside. Innovatives Supercomputing in Deutschland*, *15*(2), 128–133.

Geophysical Research Letters 10.1029/2019GL084969

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., et al. (2019). The ClimEx project: A 50-member ensemble of climate change projections at 12-km resolution over Europe and northeastern North America with the Canadian Regional Climate Model (CRCM5). *Journal of Applied Meteorology and Climatology*, *58*, 663–693. https://doi.org/10.1175/JAMC-D-18-0021.1

Lucas-Picher, P., Laprise, R., & Winger, K. (2017). Evidence of added value in North American regional climate model hindcast simulations using ever-increasing horizontal resolutions. *Climate Dynamics*, *48*(7–8), 2611–2633. https://doi.org/10.1007/s00382-016-3227-z

Mann, M. E., Rahmstorf, S., Kornhuber, K., Steinman, B. A., Miller, S. K., & Coumou, D. (2017). Influence of anthropogenic climate change on planetary wave resonance and extreme weather events. *Scientific Reports*, *7*, 45242. https://doi.org/10.1038/srep45242

Martynov, A., Laprise, R., Sushama, L., Winger, K., Separovic, L., & Dugas, B. (2013). Reanalysis-driven climate simulation over Cordex North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation. *Climate Dynamics*, *41*(11-12), 2973–3005. https://doi.org/10.1007/s00382-013-1778-9

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., et al. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1-2), 213–241. https://doi.org/10.1007/s10584-011-0156-z

Messmer, M., Gomez-Navarro, J., & Raible, C. (2015). Climatology of Vb-cyclones, physical mechanisms and their impact on extreme precipitation over central Europe. *Earth System Dynamics*, *6*(2), 541–553. https://doi.org/10.5194/esd-6-541-2015

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, *6*(4), 525–533.

Murray, R. J., & Simmonds, I. (1991). A numerical scheme for tracking cyclone centres from digital data. Part I: Development and operation of the scheme. *Australian Meteorological Magazine*, *39*, 155–166.

Muskulus, M., & Jacob, D. (2005). Tracking cyclones in regional model data: the future of Mediterranean storms. *Advances in Geosciences*, *2*, 13–19. https://doi.org/10.5194/adgeo-2-13-2005

Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., et al. (2013). IMILAST: A Community Effort to Intercompare Extratropical Cyclone Detection and Tracking Algorithms. *Bulletin of the American Meteorological Society*, *94*(4), 529–547. https://doi.org/10.1175/BAMS-D-11-00154.1

Nissen, K. M., Ulbrich, U., & Leckebusch, G. C. (2013). Vb cyclones and associated rainfall extremes over Central Europe under present day and climate change conditions. *Meteorologische Zeitschrift*, *22*(6), 649–660. https://doi.org/10.1127/0941-2948/2013/0514

Perkins, S. E., & Fischer, E. M. (2013). The usefulness of different realizations for the model evaluation of regional trends in heat waves. *Geophysical Research Letters*, *40*(21), 5793–5797. https://doi.org/10.1002/2013GL057833

Separovic, L., Alexandru, A., Laprise, R., Martynov, A., Sushama, L., Winger, K., et al. (2013). Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model. *Climate Dynamics*, *41*(11-12), 3167–3201. https://doi.org/10.1007/s00382-013-1737-5

Stahl, N., & Hofstätter, M. (2018). Vb-Zugbahnen und deren Auftreten als Serie mit Bezug zu den resultierenden Hochwassern in Bayern mit Auswirkungen auf Rückhalteräume im Isareinzugsgebiet. *Hydrologie und Wasserbewirtschaftung*, *62*(2), 78–98.

von Trentini, F., Leduc, M., & Ludwig, R. (2019). Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble. *Climate Dynamics*, *53*(3–4), 1963–1979. https://doi.org/10.1007/s00382-019-04755-8

Van Bebber, W. (1891). Die Zugstrassen der barometrischen Minima nach den Bahnenkarten der Deutschen Seewarte für den Zeitraum 1875-1890. *Meteorologische Zeitschrift*, *8*, 361–366.

Volosciuk, C., Maraun, D., Semenov, V. A., Tilinina, N., Gulev, S. K., & Latif, M. (2016). Rising Mediterranean sea surface temperatures amplify extreme summer precipitation in central Europe. *Scientific Reports*, *6*, 32450. https://doi.org/10.1038/srep32450

Willner, S. N., Levermann, A., Zhao, F., & Frieler, K. (2018). Adaptation required to preserve future high-end river flood risk at present levels. *Science Advances*, *4*(1), eaao1914. https://doi.org/10.1126/sciadv.aao1914

Zhou, Y., Hu, Q., & Wang, Y. (2018). Deep super-class learning for long-tail distributed image classification. *Pattern Recognition*, *80*, 118–128. https://doi.org/10.1016/j.patcog.2018.03.003

2.2 Paper II: A Deep Learning Approach for the Identification of Long-Duration Mixed Precipitation in Montréal (Canada)

Reference: M. Mittermeier, É. Bresson, D. Paquin & R. Ludwig (2021): A Deep Learning Approach for the Identification of Long-Duration Mixed Precipitation in Montréal (Canada), *Atmosphere-Ocean*, DOI: 10.1080/07055900.2021.1992341

Transition to paper II: Paper II addresses mixed precipitation (freezing rain and/or ice pellets) in the study area of Montréal (Québec, Canada). It is one of the costliest hydro-meteorological extreme events in Canada. Paper II uses deep learning in the form of a CNN to identify the synoptic-scale pattern that lead to pressure-driven channeling of winds along the St. Lawrence River Valley. Pressure driven channeling is the main dynamic driver of mixed precipitation events in the Montréal area, which is located about half-way along the St. Lawrence river valley. While paper I uses historical events and reanalysis data to train the neural network for atmospheric pattern detection, this study employs a large ensemble of the CRCM5 regional climate model with 18 simulations to train the CNN on a large number of more than 660,000 training examples from climate simulations. Paper II additionally applies a two-staged procedure, so that deep learning is combined with meteorological domain knowledge.

Author's contribution: All authors contributed to the study concept. All calculations, data handling and analysis were performed by MM under the supervision of EB, DP and RL. MM prepared the manuscript. It was reviewed by EB, DP and RL.

Status: published

Journal: Atmosphere - Ocean (principal scientific journal of the Canadian Meteorological and Oceanographic Society), Special Issue on: *Machine-learning Applications in the Atmospheric and Oceanic Sciences*

Impact Factor: 2.06

A Deep Learning Approach for the Identification of Long-Duration Mixed Precipitation in Montréal (Canada)

M. Mittermeier [1,*], É. Bresson[2], D. Paquin [2], and R. Ludwig [1]

[1]*Department of Geography, Ludwig-Maximilians-Universität (LMU), Munich, Germany*
[2]*Ouranos Consortium, Montréal, Quebec, Canada*

[Original manuscript received 27 January 2021; accepted 14 September 2021]

ABSTRACT *Long-duration mixed-precipitation events (freezing rain and/or ice pellets) are important cold-season hazards and understanding how climate change alters their occurrence is of high societal interest, particularly in urban areas. This study introduces a two-staged approach that employs deep learning to identify long-duration mixed precipitation over the Montréal area (Quebec, Canada) in archived climate model data using large-scale pressure patterns. The dominant dynamic mechanism leading to mixed precipitation in Montréal is pressure-driven channelling of winds along the St. Lawrence River Valley. A convolutional neural network (CNN) identifies the corresponding synoptic pattern by using a large training database derived from an ensemble of the Canadian Regional Climate Model, version 5 (CRCM5). The CRCM5 uses the diagnostic method of Bourgouin (2000) to simulate mixed precipitation and delivers training examples and corresponding class affiliations (labels) for this supervised classification task. The CNN correctly identifies more than 80% of the Bourgouin mixed-precipitation cases. In the next stage, the CNN is combined with temperature and precipitation conditions, which consider important preconditions for mixed precipitation and improve the performance of the approach. The evaluation of a CRCM5 run driven by ERA-Interim reanalysis data gives a Matthews correlation coefficient of 0.50. The deep learning approach can be applied to ensembles of regional climate models on the North American grid of the Coordinated Regional Downscaling Experiment (CORDEX-NA).*

RÉSUMÉ *[Traduit par la redaction] Les événements de précipitations mixtes de longue durée (pluie verglaçante et/ou grésil) sont des dangers importants de la saison froide et comprendre comment le changement climatique modifie leur occurrence est d'un grand intérêt sociétal, notamment dans les zones urbaines. Cette étude présente une approche en deux étapes qui utilise l'apprentissage profond pour identifier les précipitations mixtes de longue durée sur la région de Montréal (Québec, Canada) dans des données de modèles climatiques archivées en utilisant des modèles de pression à grande échelle. Le mécanisme dynamique dominant menant aux précipitations mixtes à Montréal est la canalisation des vents par la pression le long de la vallée du fleuve Saint-Laurent. Un réseau neuronal convolutif (RNC) désigne le modèle synoptique correspondant en utilisant une grande base de données d'entraînement dérivée d'un ensemble du Modèle régional canadien du climat, version 5 (MRCC5). Le MRCC5 emploie la méthode de diagnostic de Bourgouin (2000) pour simuler les précipitations mixtes et fournit des exemples d'entraînement et les affiliations de classe correspondantes (étiquettes) pour cette tâche de classification supervisée. Le RNC identifie correctement plus que 80% des cas de précipitations mixtes de Bourgouin. Dans l'étape suivante, le RNC est combiné avec les conditions de température et de précipitation, qui constituent des conditions préalables importantes pour les précipitations mixtes et améliorent le rendement de l'approche. L'évaluation d'une exécution du CRCM5 pilotée par les données de réanalyse ERA-Interim donne un coefficient de corrélation de Matthews de 0,50. L'approche d'apprentissage profond peut être appliquée à des ensembles de modèles climatiques régionaux sur la grille nord-américaine de la Coordinated Regional Downscaling Experiment (CORDEX-NA).*

KEYWORDS mixed precipitation; freezing rain; deep learning; climate pattern recognition; pressure-driven channelling; Montréal; extreme events detection

1 Introduction

Mixed precipitation in the form of freezing rain and/or ice pellets is a significant cold-season hazard in eastern North America and notably in southern Quebec, Canada. This hydrometeorological extreme event poses a great potential risk to road traffic, electrical infrastructure, and human security, particularly if it persists for several hours (McCray et al., 2019). Even though high-impact, long-duration (≥6 h) mixed-precipitation events

*Corresponding author's email: m.mittermeier@lmu.de

happen rarely, they count as one of the costliest hydrometeorological events in Canada (Dore, 2003; Ressler et al., 2012). During the extreme ice storm in 1998 in the northeastern United States and southeastern Canada, power lines were destroyed and the power supply of millions of people affected, which caused US$4 billion in damages and 47 fatalities (Cheng et al., 2004; Gyakum & Roebber, 2001; Henson et al., 2011; Milton & Bourque, 1999; Roebber & Gyakum, 2003).

Mixed precipitation occurs when a particular type of vertical temperature layering forms, where cold air ($T < 0$°C) at the surface is superposed by a relatively warm above-freezing layer ($T > 0$°C), and a cold layer ($T < 0$°C) at higher altitude. Under such circumstances freezing rain, ice pellets, or a combination of both (mixed precipitation) may form (Matte et al., 2019). Freezing rain develops through a microphysical process known as the melting process. On its way to the ground an ice crystal falls first through the warm-air layer, which is sufficiently thick that the ice crystal melts completely. The liquid drop then reaches the cold-air layer above the ground, where it becomes supercooled, which means its temperature drops below zero degrees. The supercooled liquid droplet then reaches the ground and freezes instantaneously on contact with the cold surface ($T < 0$°C; Ressler et al., 2012; Bourgouin, 2000). In contrast, the formation process of ice pellets requires a thinner warm-air layer. In this case, the droplet does not melt completely before reaching the cold-air layer. The remaining ice core facilitates rapid freezing before the droplet reaches the ground in solid form (Czys et al., 1996). Depending on the droplet size both processes can occur concurrently as mixed precipitation.

The St. Lawrence River Valley (SLRV) is a region that is especially prone to mixed precipitation because of a mechanism known as pressure-driven channelling (Ressler et al., 2012). Our study region, Montréal (see Fig. 1), is located about half-way along the SLRV, which extends from Lake Ontario to the Gulf of St. Lawrence. Under a synoptic-scale pressure pattern (≥1000 km) with higher pressure in the lower part of the SLRV (to the northeast of Montréal) and lower pressure in the upper part of the SLRV (to the southwest), cold winds from the north are channelled directly along the valley toward Montréal. These cold winds facilitate the vertical temperature structure needed for mixed precipitation by generating a cold-air layer on the surface superposed by warmer air (Ressler et al., 2012). As well as the vertical temperature structure, another necessary component for the formation of mixed precipitation is the potential for meso- or synoptic-scale ascent, which is needed for precipitation to be observed (Ressler et al., 2012). Nearly all freezing rain events in the city of Montréal are attributable to pressure-driven channelling with synoptic pressure fields being the large-scale driver (Ressler et al., 2012 Roebber & Gyakum, 2003; Stuart & Isaac, 1999;).

Understanding how climate change affects cold-season hazards, such as mixed precipitation, is an important research challenge, which can be addressed using climate models. The temporally high-resolution microphysics schemes from high-

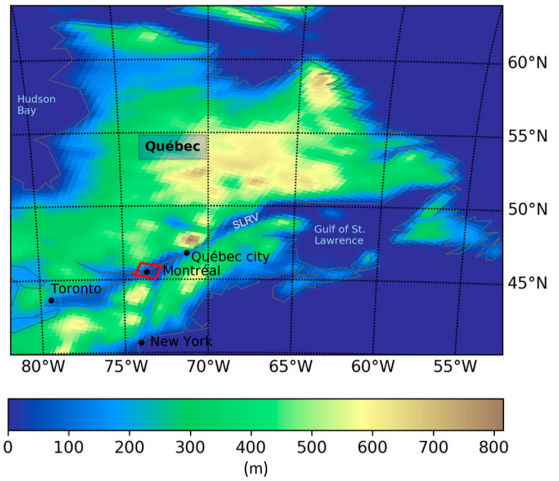

Fig. 1 Orography over Quebec at 0.11° resolution and the study area (red) covering 5 × 5 pixels centred over the city of Montréal. The Saint Lawrence River Valley (SLRV) is an important orographic feature in the process of mixed-precipitation generation.

resolution numerical weather forecasts (NWP) are too costly for global and regional climate model (RCM) simulations and can be used in convection-permitting simulations only. Therefore, several different methods that have been developed to diagnose mixed precipitation in medium resolution NWPs (e.g., Baldwin & Contorno, 1993; Cantin & Bachand, 1993; Czys et al., 1996; Ramer, 1993) are also used with RCM data, which have a resolution sufficiently high to adequately represent the necessary environmental conditions (Bresson et al., 2017; Matte et al., 2019; St-Pierre et al., 2019). The scheme of Bourgouin (2000) and its revised version (Birk et al., 2021) use the area method, which focuses on the vertical temperature profile of a layer of air and its cross-section with the 0°C isotherm. An in-line implementation of this diagnostic method is employed in the Canadian Regional Climate Model, version 5 (CRCM5; Šeparović et al., 2013; Martynov et al., 2013), which delivers the two components of mixed precipitation, freezing rain (prfr) and ice pellets (prrp), as output variables. These above-mentioned diagnostic methods require a high number of vertical pressure levels as input. By contrast, the study by Ressler et al. (2012) relates the occurrence of mixed precipitation to certain types of synoptic-scale pressure fields based on observed mixed-precipitation data analysis. A diagnostic method primarily based on the synoptic-scale dynamic drivers of mixed-precipitation events would reduce the number of input variables and levels drastically and thus allow an application on commonly available variables (e.g., from the CORDEX framework).

Artificial neural networks from the field of machine learning provide a powerful tool for visual pattern recognition and have, in some cases, successfully been applied to extreme climate pattern detection problems (Liu et al., 2016;

Mittermeier et al., 2019; Racah et al., 2017; Reichstein et al., 2019). One of the benefits of machine learning algorithms is their quick application on large datasets once they are trained. Therefore, approaches in climate sciences based on machine learning are destined for utilization with large multi-model or initial-condition ensembles (Maher et al., 2021; Mittermeier et al., 2019). The synoptic-scale pressure field related to mixed-precipitation occurrence in the SLRV due to pressure-driven channelling is an appropriate case for pattern recognition using deep learning because of the particularly well-studied connection between large-scale drivers and local extremes. In this paper, we present a new data-driven, deep learning approach to derive long-duration mixed-precipitation events over the Montréal area from large-scale pressure patterns. We describe the pressure fields according to Ressler et al. (2012) by geopotential height at 500 hPa (z500) and sea level pressure (slp) and cover a fixed domain over North America. In this case, the pattern recognition task is a binary classification task of sorting unseen synoptic patterns into two classes—mixed precipitation (class 1) or no mixed precipitation (class 0). The network is trained through a supervised training process with a training database containing training examples (image patches of the two continuous spatial variables z500 and slp at specific archived time steps) and their labels (1 or 0), both derived from CRCM5 simulations. The trained network draws inferences from the training process about the class affiliation of unseen synoptic patterns. With this, the deep neural network filters out propitious synoptic conditions for mixed precipitation.

The deep learning step (first stage) is combined with a second stage, which considers the temperature at the surface and the formation of precipitation as important preconditions for the formation of mixed precipitation (Matte et al., 2019; Ressler et al., 2012). Because the first stage is solely based on synoptic-scale dynamic drivers, the second stage ensures that important local factors are considered. As a result, the second stage enables an improvement in the precision of the deep learning approach. High-resolution temperature and total precipitation data from the CRCM5 are used to check whether the air temperature and the precipitation rate at the surface lie below and above a certain threshold, respectively.

Besides the low number of required input variables and levels, an additional benefit of this method is its quick computation because deep learning is designed to be used with large datasets. Thus, this deep learning approach can be utilized with large multi-model or initial-condition large ensembles.

This study introduces a two-staged approach that uses deep learning for the identification of long-duration mixed precipitation in archived climate model data for the Montréal study area. Our work is based on Ressler et al. (2012) and Roebber and Gyakum (2003), who relate the mesoscale process of channelling along the SLRV to synoptic-scale pressure fields. Thus, this study focuses on the large-scale drivers of an extreme event and captures them through pressure fields over North America. Using this case study, we highlight a challenge of deep learning for spatiotemporal pattern recognition regarding the test set choice and suggest a best practice for such applications in climate science.

2 Data

In order to train a deep neural network (DNN) on the identification of the synoptic-scale pressure pattern related to long-duration mixed precipitation in the Montréal area, a large training database is required. Acquiring a large number of labelled training examples with known class affiliations is a common challenge in deep learning applications in earth system sciences (Reichstein et al., 2019). We employ an ensemble of CRCM5 simulations, which provides 1437 years of labelled data (see Table 1), archived at a three hourly temporal resolution for the variables slp, z500, prfr, and prrp. The variables slp and z500 are chosen to describe the synoptic-scale pressure patterns because their relation to the occurrence of mixed precipitation in Montréal has been well studied (Ressler et al., 2012). The CRCM5 simulations cover the North American domain of the Coordinated Regional Downscaling Experiment (CORDEX-NA; Mearns et al., 2017) on a rotated-pole grid and have a spatial resolution of 0.22 (CORDEX domain name: NAM-22)

The CRCM5 simulations are driven by four different general circulation models (GCMs): the second generation Canadian Earth System Model (CanESM2; Arora et al., 2011), the Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5 (CNRM-CM5; Voldoire et al., 2013), the Max Planck Institute for Meteorology-Earth System Model-Low Resolution (MPI-ESM-LR; Stevens et al., 2013), and the Geophysical Fluid Dynamics Laboratory-Earth System Model, version 2M (GFDL-ESM2M; Dunne et al., 2012). Observed greenhouse gas

TABLE 1. Overview of the CRCM5 ensemble used as a training database. The CRCM5 simulations are driven by four different GCMs and three different GHG concentration scenarios. One simulation is driven by the ERA-Interim reanalysis (ID 19). In case several members of a model exist, the number is indicated by -m plus member number.

ID	RCM	GCM	Time series	GHG concentrations
1	CRCM5	CanESM2 -m1	1950–2005	observed
2			2006–2100	RCP8.5
3				RCP4.5
4	CRCM5	CanESM2 -m3	1950–2005	observed
5			2006–2100	RCP8.5
6	CRCM5	CanESM2 -m4	1950–2005	observed
7			2006–2100	RCP8.5
8	CRCM5	CanESM2 -m5	1950–2005	observed
9			2006–2100	RCP8.5
10	CRCM5	CNRM-CM5 -m1	1950–2005	observed
11			2006–2100	RCP8.5
12				RCP4.5
13	CRCM5	MPI-ESM-LR -m1	1950–2005	observed
14			2006–2100	RCP8.5
15				RCP4.5
16	CRCM5	GFDL-ESM2M -m1	1950–2005	observed
17			2006–2100	RCP8.5
18				RCP4.5
19	CRCM5	ERA-Interim	1979–2017	observed

(GHG) concentrations are used from 1950 to 2005 (seven simulations). From 2006 on the representative concentration pathways (RCP8.5: seven simulations; RCP4.5: four simulations; see Table 1) are used as future projection scenarios. Because synoptic-scale patterns are in large part dominated by the driving GCM, using only one GCM would result in an unwanted specific bias. Using historic periods as well as scenarios with high GHG concentrations means that the training set covers simulations with a range of different severities of climate change. For CanESM2, four members are available with RCP8.5, for which the initial conditions were slightly altered in order to simulate internal variability. Even though only one RCM is used, we provide a good variety of training examples with this ensemble of different GCMs, different scenarios, and different initial-condition members.

One additional simulation is driven by the ERA-Interim reanalysis (CRCM5–ERA-Interim run; Dee et al., 2011) and is used for evaluation purposes.

3 Methodology

This section describes the two stages of our deep learning approach: the first stage, in which the convolutional neural network (CNN) is employed, and the second stage, which covers further processing of the CNN output with a temperature condition and a precipitation condition.

a *A First Stage: Convolutional Neural Network (CNN)*

The workflow of the training set preparation is illustrated in Fig. 2.

1 Event Definition For Label Preparation

Every archived cold-season time step of the CRCM5 ensemble is a potential training example for the supervised deep learning approach, either for class 1 (mixed precipitation) or class 0 (no mixed precipitation). To derive the label information for a certain archived time step from CRCM5 Bourgouin in-line, mixed-precipitation events are defined as follows: mixed precipitation occurs whenever the sum of freezing rain and ice pellets at any pixel in the study area exceeds a threshold of 0.125 mm per three hours (with reference to 1 mm per day) ($prfr + prrp > 0.125$ mm; Lambert & Hansen, 2011; Matte et al., 2019). St-Pierre et al. (2019) state that the Bourgouin (2000) scheme implemented in-line in the CRCM5 reproduces freezing rain in Montréal best at a spatial model resolution of 0.22°, which is the resolution of the CRCM5 ensemble used in this study. Only the cold season of October to April is considered (Henson & Stewart, 2007). The study area covers a 5×5 grid over the Montréal area (42.6° to 43.3°N, −108.3° to −107.3°E) and is illustrated in Fig. 1. In accordance with Ressler et al. (2012) the relationship between large-scale drivers and mixed-precipitation occurrence is only applied to long-duration events. Short-duration events, which last less than six hours (Matte et al., 2019; Ressler et al., 2012), are neglected. The long-duration events (≥ 6 h) also include near-continuous events, which may include fewer than six hours of non-mixed precipitation within the course of the event as long as the condition of at least six hours of continuous precipitation is fulfilled (Ressler et al., 2012). In the next step, synoptically dependent events are removed. Synoptical independence is defined as at least three days of non-mixed precipitation between two mixed-precipitation events (Ressler et al., 2012).

2 Setup Division

The cold-season, long-duration mixed-precipitation events resulting from the event definition provide training examples for class 1. The class 1 training examples are split into three subsamples: 90% of events are used for training, 5% for development, and 5% for testing (see Fig. 2). The training subset is used to train the DNN. The development set is employed during the training process to tune the hyperparameters within the network. This subset is also used to monitor the issue of overfitting (e.g., through learning curves). The test set is used solely to evaluate the network performance after the training process has been completed and remains untouched before that. The events, which have differing durations (on average 9.3 h) are treated as entities for the setup division in order to avoid single time steps of one and the same event so that similar synoptic-scale pressure patterns are distributed over all three training subsamples. The proportion of 90%, 5%, 5% for the setup division is chosen because of our high number of class 1 training examples (65,970), which allows us to use a higher percentage for the training subset. Additionally, the higher ratios of class 1: class 0 in the development and test subsets compared with the training subset (see Section 3.a.3) influenced the setup division because this leads to a high number of counter-examples in the development and test subsets. In summary, the training subset covers 661,200 training examples with 415,261 in the training subset, 121,064 in the developement set, and 124,875 in the test set.

3 Resolving The Imbalanced Class Distribution

A crucial point in building up the training set is dealing with the imbalanced class distribution, a challenge intrinsic to all classification tasks dealing with extreme events. Mixed-precipitation time steps in class 1 make up only 2.7% of all cold-season time steps averaged over the 18 simulations of the CRCM5 ensemble. This corresponds to a ratio of 1:36 between positive examples (class 1) and negative examples (class 0). The CRCM5-ERA-Interim run used for evaluation purposes has a ratio of 1:31. To avoid the network overlooking the minority class during training, undersampling is used. Therefore, the ratio in the training set is set to 1:6, while in the development and test sets a target ratio of 1:36 is used. From all potential counter-examples in the CRCM5 ensemble as many class 0 training examples are selected as match with the corresponding ratio. The choice of counter-examples is carried out randomly in order to ensure high diversity. A rule for synoptic independence from mixed-precipitation events is introduced, which means that a three-day buffer is set around all mixed-precipitation events, so that no

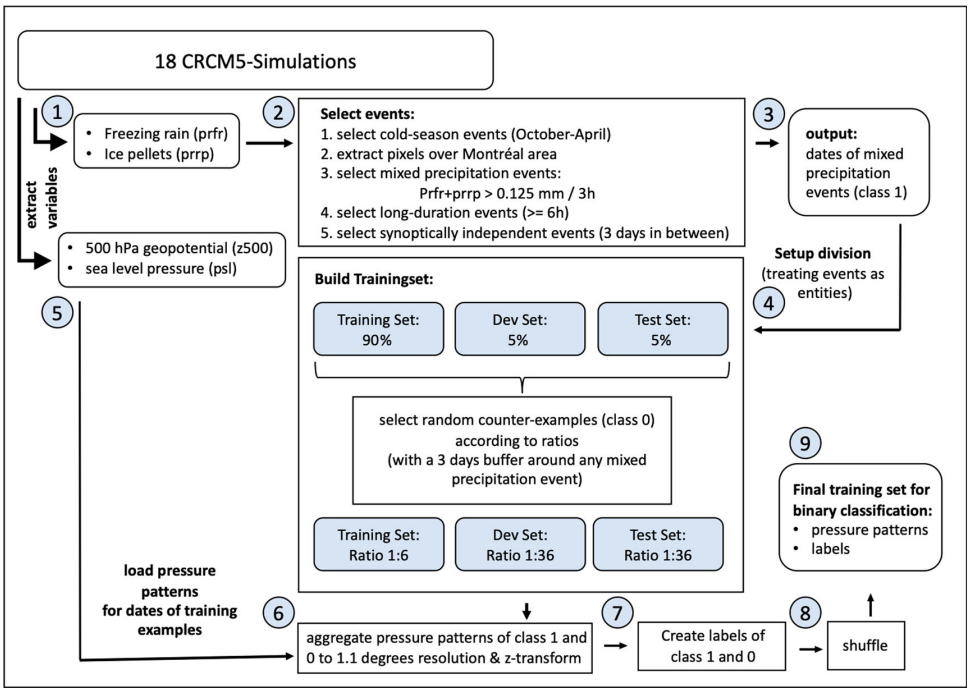

Fig. 2 Workflow graphic showing the training set preparation including event definition and setup division.

counter-examples are chosen from the period shortly before or shortly after a mixed-precipitation event. This avoids the network being trained on arbitrary transition phases. A buffer between counter-examples, though, is not possible because 25% of all non-mixed-precipitation cold-season time steps in the CRCM5 ensemble are already used in the training database. Putting a buffer of, for example, three days around every counter-example would by far exceed the number of available time steps. As a consequence, one counter-example in the test set might be the following archived time step to a counter-example in the training subset.

In the end, classes 1 and 0 are combined and shuffled so that the training examples are in no particular chronological order.

4 Aggregation To A Coarser Resolution

The atmospheric pressure information (z500 and slp) is pre-processed for every training example. This pre-processing includes an aggregation of the original resolution of the CRCM5 simulations of 0.22° (300×340 pixels over CORDEX-NA) to a coarser resolution. Because of the smooth character of the synoptic fields, a high spatial resolution is not necessary to describe the characteristics of the pressure patterns, whereas a high spatial resolution is needed in the second stage when using the temperature and precipitation conditions (St-Pierre et al., 2019). However, a high number of pixels, and thus nodes, makes the neural network more complex and more vulnerable to overfitting. For this reason, the training data are aggregated to a spatial resolution of 1.1° (60×68 pixels over CORDEX-NA) by calculating the mean of windows of 5×5 pixels.

5 Standardization

The training data are standardized using a z-transformation. A z-transformation is carried out by subtracting the mean of the training subset from each training example and dividing it by the variance of the training subset. This corresponds to pressure anomalies. Figure 3 shows the pressure anomaly patterns averaged over all training examples of class 1 (training, development, and test sets). For both variables (slp and z500) a distinct low-pressure system in the northern part of the SLRV and a high-pressure system in the southern part is visible, which is slightly shifted away from the SLRV axis at higher altitudes (z500).

6 Description Of A Cnn

In the era of deep learning, CNNs are a widespread state-of-the-art network type in the field of computer vision and visual pattern recognition (LeCun et al., 2015). The deep architecture of CNNs includes several convolutional and pooling layers before one or more fully connected layers. The inputs to a CNN are the number of training examples (m). Each training example is an image of the shape (nH, nW, nC) with nH and nW being the height and width of the image in pixels and nC the number of channels (Liu et al., 2016). Here, there are two channels, one for each variable (z500 and slp). As output, the CNN gives a vector of shape (m, nY) with nY being the number of classes. The output

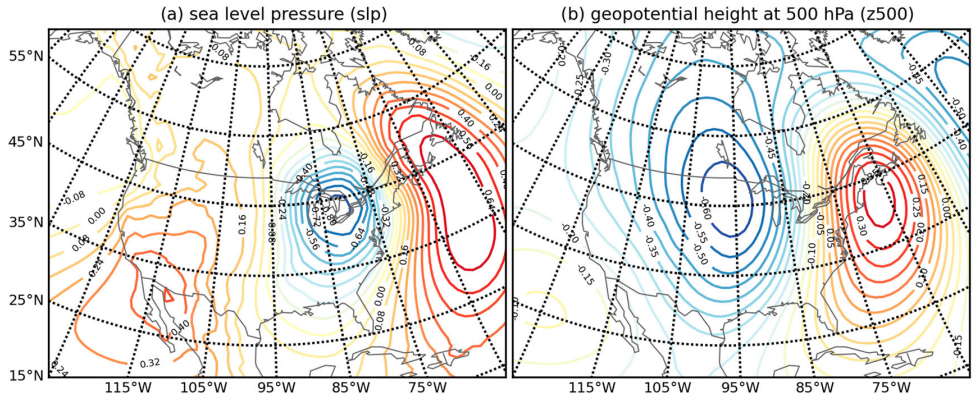

Fig. 3 Mean pressure anomaly patterns during long-duration mixed-precipitation events in the Montréal area for the variables (a) slp and (b) z500. The plots are generated by calculating the average of all standardized training examples of class 1. The colours visualize the values indicated for the contours with a colour range from blue (low values) to red (high values).

indicates the predicted class affiliation for every training example and every class category, in this case two (1,0). In the convolutional layers a number of k filters (kernels) are applied to the input image or the input feature map (output of a convolutional layer). Such filters are designed for edge detection and feature extraction and break down the complexity of the input pattern into more and smaller features (LeCun et al., 2015). During this step, the image size (nH, nW) is reduced while the number of channels (nC) remains the same. The output of the convolutional layer is k feature maps. In a subsequent pooling layer further filters are applied to the feature maps. Within each filter the maximum or mean value (max or mean pooling) of the feature map's pixel values are calculated. During this step, the size of the feature maps is further reduced. In the first fully connected layer the shape of the feature maps is flattened to a vector of nodes with length (nH, xn, Wx, nCx, k; Liu et al., 2016). In the following fully connected layers each node is connected with all nodes in the previous layer until in the output layer the final class predictions are generated. The learning process of a CNN is typically conducted using a backpropagation algorithm (LeCun et al., 2015). During backpropagation, a cost function, which compares the predicted class scores to the correct label information, is gradually minimized through a process known as gradient descent. The error of the cost function is backpropagated through the entire network for each iteration; the internal weight parameters of the network are adjusted, and the entire processing of the network (feed-forward propagation) is then repeated until a local minimum of the cost function is found (Liu et al., 2016).

TABLE 2. CNN architecture with seven layers and their layer specifications based on Liu et al. (2016).

Layer Type	Layer Specifications
Conv2D	window size: 5×5, #filters = 8, stride = 1
MaxPooling2D	window size: 2×2, stride = 2
Conv2D	window size: 5×5, #filters = 16, stride = 1
MaxPooling2D	window size: 2×2, stride = 2
Dense	50 nodes
Dropout	r-term: 0.9
Dense	nodes: 2

7 Network Configuration

A binary cross entropy function is chosen as the cost function because it is optimized for binary classification tasks (Kline & Berardi, 2005). The backpropagation is carried out for batches of 64 training examples instead of for all 661,200 training examples at once (batch gradient descent). The number of epochs chosen is 20, which means that the iterations are continued until all training examples are shown to the network 20 times. A rectified linear units function (RELU) is applied as an activation function in all layers except the output layer, where the sigmoid activation generates class scores in the range of [0,1].

8 Network Architecture

Table 2 shows the network architecture of the CNN. The architecture is chosen according to Liu et al. (2016), who designed it for similar image sizes. The shallow architecture with only seven layers has the advantage of reducing the risk of overfitting. Furthermore, the large kernel (filter) sizes are adapted for climate patterns with relatively simple patterns compared with applications on pictures, where smaller kernel sizes are common (Liu et al., 2016). The network architecture consists of two sequences of a convolutional layer (Conv2D) and a subsequent max-pooling layer (MaxPooling2D). In the convolutional layer, kernels (filters) of 5×5 pixels are applied to the input image or feature map. The number of filters is eight in the first convolutional layer and 16 in the second convolutional layer. The stride is one for both cases, which means the kernels are shifted by one pixel while they slide over the input image or feature map. Subsequent to the second MaxPooling2D layer a fully connected, dense neural network layer with 50 nodes follows. Afterwards, a dropout layer is incorporated for

regularization purposes. With dropout, some neurons are deactivated during training in order to avoid the risk of overfitting. The rate of the dropout layer (r-term = 0.9), as well as the learning rate for backpropagation ($\alpha = 0.001$) are determined by hyperparameter tuning using a random search of the Keras-Tuner library. With random search, different combinations of the hyperparameters are tested and the combination with the highest accuracy on the development set is determined. The output layer is a second dense layer with two nodes for the binary classification task.

In addition to the undersampling strategy in dealing with the imbalanced class distribution, error-weighting with a factor of six (corresponding to the ratio in the training set) is applied during training. This way, an error the network makes on the minority class is weighted more strongly than an error on the majority class.

9 Post-Processing: Removing Short-Duration Events

The CNN is exclusively trained on long-duration events with durations of at least six hours. The network, however, affiliates every individual input pattern of a time series with one of the two classes, independent of the class affiliation of the previous archived time step. This is why the direct output of the network also contains single short-duration events. These events are modified to belong to class 0 in a post-processing step because the network is not designed for the identification of short-duration mixed-precipitation events. Therefore, the timeline of the test set is sorted in chronological order. Single positive network predictions are removed if both archived (3-hourly) time steps before and after the analyzed time step are classified as negative. This way, near-continuous events with a single time step of non-mixed precipitation in between are still considered.

b *Second Stage: Temperature and Precipitation Conditions*

1 Temperature Condition

In the second stage, after applying the CNN to a target time series, the positive network predictions undergo an additional procedure and are checked for fulfilling the temperature condition. This step checks whether the temperature, as an important thermodynamic condition for the occurrence of mixed precipitation, is below a certain threshold associated with the formation of mixed precipitation and thereby corrects some false positives. The temperature condition is passed whenever at least one pixel in the study area has a temperature below 0°C (Bourgouin, 2000; Liang & Sushama, 2019; McCray et al., 2020; St-Pierre et al., 2019).

The temperature criterion is applied to the 2 m air temperature (tas) of the high-resolution RCM data over the study area of Montréal.

2 Precipitation Condition

So far, our method identifies the synoptic drivers of mixed precipitation in Montréal and considers an important thermal condition for the formation of mixed precipitation through the temperature condition. In the next and last step of the second stage, all positive network predictions are further investigated with respect to the occurrence of any kind of precipitation (liquid, mixed, or solid). Only if the variable total precipitation (pr) indicates a positive value do the positive network predictions pass. The precipitation condition is fulfilled whenever the maximum total precipitation of the pixels in the study area is greater than 0.125 mm per three hours, which is the same threshold used for the event definition in Section 3.a.1. In the same manner as with the temperature criterion, the precipitation condition is applied to the high-resolution RCM data. In the case of CRCM5 the spatial resolution is 0.22°, summing up to 25 pixels over the study area. For an application of the deep learning approach to other RCMs the spatial resolution available for the temperature and precipitation conditions might differ.

4 Results

The performance results of the deep learning approach are presented in two sections. In the first part the classification results on the training database, the CRCM5 ensemble, are presented in comparison with the method in Bourgouin (2000). In the second part an evaluation is carried out on an additional, unseen time series, the CRCM5–ERA-Interim run, to illustrate the robustness of the approach under reference conditions.

a *Classification Results on the Training Database*

1 Network Performance

The CNN correctly classifies 93.3% of all cases in the test set. This means the network affiliates 93.3% of all archived time steps in the CRCM5 ensemble to the same class as the labels derived from the Bourgouin scheme implemented in-line in the CRCM5. On the training subset it performs with an overall accuracy of 93.9% and on the development set with an accuracy of 93.5%. The small differences between the subsets indicate that the network generalizes well from the training subset to the other two subsets and does not overfit to the training subset. After removing the short-duration events from the network predictions, the overall accuracy on the test set improves to 95.7%.

The learning curves in Fig. 4 denote the development of the loss (Fig. 4a) and accuracy (Fig. 4b) on the training and the development set during the 20 training epochs. At the end of the training process the two graphs are close to overlapping. Although the accuracy increases steadily for the training subset, there are more fluctuations for the development set, which tends to decrease toward the end of the training process. This is why the training is stopped at 20 epochs.

2 Classification Results Of The CNN (First Stage)

The overall accuracy does not consider the imbalanced class distribution of the target data and thus neglects the accuracy paradox (Chicco & Jurman, 2020; Valverde-Albacete & Peláez-Moreno, 2014). Because of this, further performance

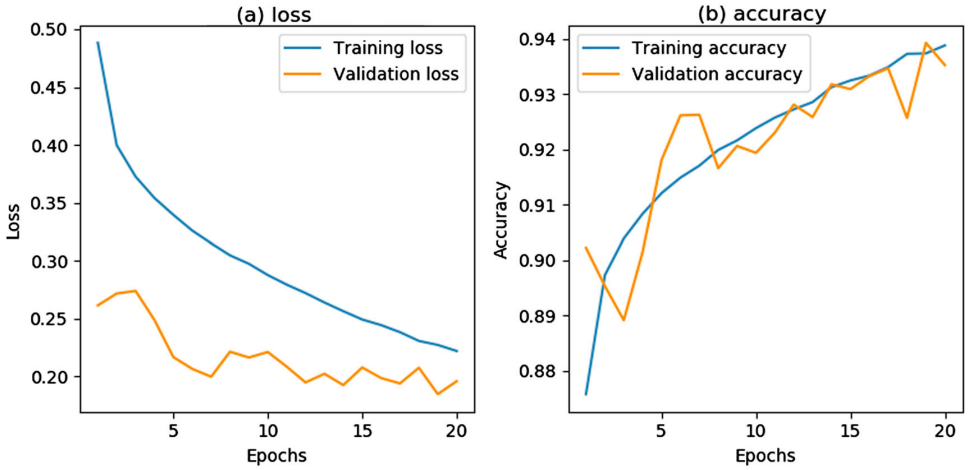

Fig. 4 Learning curves with the (a) loss and (b) accuracies of the training and the development set during the 20 training epochs.

measures are chosen: confusion matrices with recall and precision; the F1 score, which is a single-number metric that combines precision and recall and has a range of [0,1]; and the Matthews correlation coefficient (MCC; Chicco & Jurman, 2020; Matthews, 1975). An advantage of the MCC over the F1 score is that it considers the true negatives and additionally assesses the network performance compared with random guessing. The range of the MCC index is [-1,1] with values close to zero showing no improvement compared with random guessing (Chicco & Jurman, 2020).

The confusion matrix in Table 3 shows the classification results of the CNN after removing short-duration events. The recall indicates that the CNN correctly identifies 89.9% of all mixed-precipitation time steps that are also identified by the Bourgouin (2000) scheme, which is used as a baseline in this study. The precision denotes that 36.6% of all identified mixed-precipitation time steps are indeed classified as mixed precipitation according to Bourgouin. The F1 score has a value of 0.52. The confusion matrix further shows the number of true and false predictions. The number of false positives, also referred to as a type I error, is 5058. The number of type II errors, false negatives, is 327. The MCC index has a value of 0.56.

TABLE 3. Confusion matrix with archived time steps (3-hourly) of the test set (first stage). The columns show the labels derived from the CRCM5 run (Bourgouin, 2000), and the rows show the network predictions. The right column shows the precision (*P*) and the negative predictive value (NPV). The last row shows the recall (*R*) and the false positive rate (FPR). The lowest right cell displays the overall accuracy (*A*).

		Labels		(%)
		Class 1	Class 0	
Predictions	Class 1	2922	5058	*P* = 36.6
	Class 0	327	116568	NPV = 99.7
(%)		*R* = 89.9	100-FPR = 95.8	*A* = 95.7

3 Classification Results After Applying The Temperature And Precipitation Conditions (Second Stage)

The overall accuracy after applying the temperature and precipitation conditions (the second stage of the two-staged approach) is 98.8% on the test set (see Table 4). The recall is slightly affected by the second stage and decreases, mainly due to the temperature condition, to 83.8%. The precision, on the other hand, improves significantly by 35.8% to a value of 72.4% because 4020 false positives are corrected to true negatives by the temperature and precipitation conditions. As a result, the F1 score and MCC both improve to 0.78, which is an increase of 21% in case of the MCC.

TABLE 4. Confusion matrix of the test set after applying the temperature and precipitation conditions (second stage). See Table 3 for a full description.

		Labels		(%)
		Class 1	Class 0	
Predictions	Class 1	2721	1038	*P* = 72.4
	Class 0	528	120588	NPV = 99.6
(%)		*R* = 83.8	100-FPR = 99.2	*A* = 98.8

b *Evaluation on the CRCM5–ERA-Interim Run*

1 Classification Results On The CRCM5–ERA-Interim Run (First Stage)

The classification results on the additional time series of the CRCM5–ERA-Interim run are significantly lower than the classification results on the test set (see Table 5). The overall accuracy is 87.5%, the recall 87.2%, and the precision 18.6%. The F1 score has a value of 0.31. The MCC is 0.37.

2 Classification Results On CRCM5–ERA-Interim Run After Applying The Temperature And Precipitation Conditions (Second Stage)

The overall accuracy improves to 94.0% after applying the temperature and precipitation conditions (second stage). The recall decreases to 82.6% (see Table 6). The precision increases to 32.4%. Therefore, the second stage of the deep learning approach increases the precision by only 13.8% for the CRCM5–ERA-Interim run. This is a considerably low increase compared with the improvement of 35.8% in the test set. The F1 score is 0.47, and the MCC 0.50.

5 Discussion

This study introduces a two-staged approach that combines deep learning techniques with traditional conditions for the identification of the synoptic pattern related to pressure-driven channelling at the SLRV, the dominant dynamic mechanism leading to long-duration mixed precipitation in the study area of Montréal. Our approach takes the mixed-precipitation events simulated by CRCM5 using the Bourgouin (2000) in-line diagnostic method as the baseline and thus as "ground truth." This ground truth has been evaluated with observations by Bresson et al. (2017) and St-Pierre et al. (2019). The CRCM5 using the Bourgouin in-line implementation was chosen as the baseline in order to have a large amount of training data for the CNN. The evaluation of the Bourgouin scheme and of its use in the CRCM5 (Bresson et al., 2017; St-Pierre et al., 2019) is not within the scope of this study.

a *A Summary of Performance on Test Set*

A deep CNN is trained on a large training database derived from an ensemble of CRCM5 simulations driven by different GCMs following observed and two GHG concentration scenarios and includes internal variability experiments. The resulting algorithm (first stage of the approach) has a good recall of 90% on the test set. The drawback is the low precision of 37%. In order to consider important preconditions for the formation of mixed precipitation, a temperature condition and a precipitation condition are applied to the high-resolution input data in the second stage of the approach. This reduces the type I error significantly and increases the precision by 35.8%, while the recall is only slightly affected. The final MCC value on the test set is 0.77.

b *Discrepancy in Performance Between Test Set and CRCM5–ERA-Interim*

The evaluation of the deep learning approach on an unseen time series, the CRCM5–ERA-Interim run, shows that the temperature and precipitation conditions have a lower effect than on the test set. The difference is largest for the temperature condition, with an improvement in the precision of +5.7% for the evaluation set compared with +19.7% for the test set. This means that a higher percentage of false positives is corrected in the case of the test set from CRCM5 simulations than for the CRCM5–ERA-Interim runs. This is probably because an absolute temperature threshold (0°C) is used, which is exceeded more often as global warming progresses. Consequently, the criterion is correctly violated more often in the simulations of a future climate with strong temperature increases than in the CRCM5–ERA-Interim runs, which covers the historical period 1979–2017. The dependence of the temperature criterion on a single absolute threshold could be reduced by considering the entire vertical temperature profile, which we discuss in Section 5.d

Also, the evaluation with CRCM5–ERA-Interim exhibits lower performance measures compared with the test set. The MCC after the second stage is 0.50 while it is 0.77 for the test set. This discrepancy likely stems from the choice of test set because the counter-examples in the test set are not entirely independent of the counter-examples in the training subset. However, the setup division does not allow an implementation of a buffer between counter-examples, which would ensure independence among them. This raises doubts about the reliability of the precision measurement on the test set. For precision, the F1 score and MCC, it is assumed that the performance values of the evaluation are more reliable than the ones of the test set. The performance measure of class 1 is not affected by this issue, and the recall is with 82.6% (CRCM5–ERA-Interim) and 83.8% (test set) very stable between the different datasets.

c *Recommendations for Test Set Choice*

For future work it is recommended that an entire time series be separated from the training process in order to use it for the development and test set (see e.g., Racah et al., 2017). In our approach one of the CRCM5–CanESM2 members could be used for evaluation purposes only. While doing so, it is important to ensure that the ratio in the development

TABLE 5. Confusion matrix on the CRCM5–ERA-Interim run (first stage). See Table 3 for a full description.

		Labels		(%)
		Class 1	Class 0	
Predictions	Class 1	1787	7798	$P = 18.6$
	Class 0	262	54681	NPV = 99.5
(%)		$R = 87.2$	100-FPR = 87.5	$A = 87.5$

TABLE 6. Confusion matrix on the CRCM5–ERA-Interim run after applying the temperature and precipitation conditions (second stage). See Table 3 for a full description.

		Labels		(%)
		Class 1	Class 0	
Predictions	Class 1	1693	3526	$P = 32.4$
	Class 0	356	58953	NPV = 99.4
(%)		$R = 82.6$	100-FPR = 94.4	$A = 94.0$

and test set corresponds to the expected ratio in the target data, the data the CNN is meant to be applied to in the end.

It must be stated that the issue of choosing a test set is critical for climate extreme event detection tasks that deal with time series of climate data and strongly imbalanced datasets. Presenting the test set composition and examining the reliability of the test set is a key point for high transparency in machine learning applications and it should be best practice to present the test set composition, especially when dealing with spatiotemporal data as is common in climate sciences.

d *Benefits and Limitations of the Two-Staged Approach Using Deep Learning*

The two-staged approach introduced in this study can be applied to archived climate model data of different RCMs as long as they can reasonably be aggregated to a resolution of 1.1° on the CORDEX-NA domain. For the second stage a high resolution of 0.22° is recommended.

The main constraint of using deep learning is the moderate precision of 32.4% for CRCM5–ERA-Interim. One strategy for future work would be to focus performance improvement efforts on the choice of the input variables and to compare the use of different geopotential height fields as predictors at the beginning of the deep learning loop because these are essential for the entire approach. For example, the vertically integrated variable slp could be replaced or complemented by geopotential height at, for example, 850 or 925 hPa, the major benefit of which would be the inclusion of information about the vertical structure. The variable choice in this study was influenced by the goal of making our approach applicable to commonly available variables (e.g., from the CORDEX framework). Future datasets that have more available variables will allow different predictor choices in future work while still ensuring the broad transferability of the approach. In line with this, a further strategy for potential performance improvement is the inclusion of the entire local vertical temperature profile over the study area of Montréal. This would allow a better representation of the local physical processes that are needed for the generation of mixed precipitation. With future datasets that provide more vertical temperature levels, the temperature criterion in the second stage could be expanded and improved so that it not only considers the near-surface temperature (tas) but also utilizes more temperature levels for a better description of the vertical temperature profile.

Further constraints of our approach are its limitation to long-duration events (≥6 h) and its spatial limitation to the Montréal area.

One of the major benefits of the deep learning approach is that it is based on physical, dynamic mechanisms because the synoptic-scale pressure field is used, which leads to pressure-driven channelling of winds over the SLRV and mixed precipitation in the study area; therefore, results remain explainable. The diagnostic scheme of Bourgouin (2000), employed for the training dataset, is also physically based. Furthermore, the deep learning approach only needs slp, z500, tas, and pr as input variables. Despite pre-processing of the data, which includes a potentially necessary transformation to the desired grid and the z-transformation, the application of the DNN takes little computing time (41 seconds for the CRCM5–ERA-Interim run corresponding to 1.1 seconds per year; 1 node, 48 cores) and is appropriate for application to large RCM ensembles.

Acknowledgements

The trained CNN presented in this paper plus the PYTHON code for its application is freely available via zenodo (https://doi.org/10.5281/zenodo.4404600 and https://doi.org/10.5281/zenodo.5495623). The calculations were carried out using tensorflow and keras in Python. The CRCM5 data were generated and supplied by Ouranos. The CRCM5 was developed at the Université du Québec à Montréal (UQAM) in collaboration with Environment and Climate Change Canada (ECCC). The CRCM5 computations were made on the supercomputer Guillimin at McGill University, which is managed by Calcul Québec and Compute Canada. The operation of the supercomputer is funded by the Canada Foundation for Innovation (CFI), the Ministère de l'Économie, de la Science et de l'Innovation du Québec (MESI), and the Fonds de recherche du Québec – Nature et technologies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

M. Mittermeier http://orcid.org/0000-0002-8668-281X
D. Paquin http://orcid.org/0000-0002-1353-930X
R. Ludwig http://orcid.org/0000-0002-4225-4098

References

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G., & Merryfield, W. J. (2011). Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, *38*(5), L05805. https://doi.org/10.1029/2010GL046270

Baldwin, E. B., & Contorno, S. P. (1993, 2–3 August). *Development of a weather-type prediction system for NMC's mesoscale ETA model.* Preprints 13th conference on Weather analysis and forecasting (pp. 86–87), Vienna, VA, United States. American Meteorological Society.

Birk, K., Lenning, E., Donofrio, K., & Friedlein, M. T. (2021). A revised Bourgouin precipitation-type algorithm. *Weather and Forecasting*, *36*(2), 425–438. https://doi.org/10.1175/WAF-D-20-0118.1

Bourgouin, P. (2000). A method to determine precipitation types. *Weather and Forecasting*, *15*(5), 583–592. https://doi.org/10.1175/1520-0434(2000)015%3C0583:AMTDPT%3E2.0.CO;2

Bresson, É, Laprise, R., Paquin, D., Thériault, J. M., & de Elía, R. (2017). Evaluating the ability of CRCM5 to simulate mixed precipitation. *Atmosphere-Ocean*, *55*(2), 79–93. https://doi.org/10.1080/07055900.2017.1310084

Cantin, A., & Bachand, D. (1993). S*ynoptic pattern recognition and partial thickness techniques as a tool for precipitation types forecasting associated with a winter storm*. Service de l'environnement atmospherique, Region du Quebec, Centre météorologique du Québec.

Cheng, C. S., Auld, H., Li, G., Klaassen, J., Tugwood, B., & Li, Q. (2004). An automated synoptic typing procedure to predict freezing rain: An application to Ottawa, Ontario. *Weather and Forecasting*, *19*(4), 751–768. https://doi.org/10.1175/1520-0434(2004)019%3C0751:AASTPT%3E2.0.CO;2

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), 6. https://doi.org/10.1186/s12864-019-6413-7

Czys, R. R., Scott, R. W., Tang, K. C., Przybylinski, R. W., & Sabones, M. E. (1996). A physically based, nondimensional parameter for discriminating between locations of freezing rain and ice pellets. *Weather and Forecasting*, *11*(4), 591–598. https://doi.org/10.1175/1520-0434(1996)011%3C0591:APBNPF%3E2.0.CO;2

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., … Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Dore, M. H. I. (2003). Forecasting the conditional probabilities of natural disasters in Canada as a guide of disaster preparedness. *Natural Hazards*, *28* (2/3), 249–269. https://doi.org/10.1023/A:1022978024522

Dunne, J. P., John, J., Adcroft, A., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W. F., Dunne, K. A., Harrison, M. J., Krasting, J. P., Malyshev, S., Milly, P. C. D., Phillipps, P., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A. T., & Zadeh, N. (2012). GFDL's ESM2 global coupled climate-carbon Earth System models. Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, *25*(19), 6646–6665. https://doi.org/10.1175/JCLI-D-11-00560.1

Gyakum, J. R., & Roebber, P. J. (2001). The 1998 ice storm - analysis of a planetary-scale event. *Monthly Weather Review*, *129*(12), 2983–2997. https://doi.org/10.1175/1520-0493(2001)129%3C2983:TISAOA%3E2.0.CO;2

Henson, W., & Stewart, R. (2007). Severity and return periods of icing events in the Montréal area. *Atmospheric Research*, *84*(3), 242–249. https://doi.org/10.1016/j.atmosres.2006.08.002

Henson, W., Stewart, R., Kochtubajda, B., & Thériault, J. (2011). The 1998 ice storm: Local flow fields and linkages to precipitation. *Atmospheric Research*, *101*(4), 852–862. https://doi.org/10.1016/j.atmosres.2011.05.014

Kline, D. M., & Berardi, V. L. (2005). Revisiting squared-error and cross-entropy functions for training neural network classifiers. *Neural Computing and Applications*, *14*(4), 310–318. https://doi.org/10.1007/s00521-005-0467-y

Lambert, S. J., & Hansen, B. K. (2011). Simulated changes in the freezing rain climatology of North America under global warming using a coupled climate model. *Atmosphere-Ocean*, *49*(3), 289–295. https://doi.org/10.1080/07055900.2011.607492

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521* (7553), 436–444. https://doi.org/10.1038/nature14539

Liang, J., & Sushama, L. (2019). Freezing rain events related to atmospheric rivers and associated mechanisms for western North America. *Geophysical Research Letters*, *46*(17-18), 10541–10550. https://doi.org/10.1029/2019GL084647

Liu, Y., Racah, E., Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., & Collins, W. (2016). Application of deep convolutional neural networks for detecting extreme weather in climate datasets. https://arXiv.org/abs/1605.01156.

Maher, N., Milinski, S., & Ludwig, R. (2021). Large ensemble climate model simulations: Introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth System Dynamics*, *12*(2), 401–418. https://doi.org/10.5194/esd-12-401-2021

Martynov, A., Laprise, R., Sushama, L., Winger, K., Šeparović, L., & Dugas, B. (2013). Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: Model performance evaluation. *Climate Dynamics*, *41*(11-12), 2973–3005. https://doi.org/10.1007/s00382-013-1778-9

Matte, D., Thériault, J. M., & Laprise, R. (2019). Mixed precipitation occurrences over southern Québec, Canada, under warmer climate conditions using a regional climate model. *Climate Dynamics*, *53*(1-2), 1125–1141. https://doi.org/10.1007/s00382-018-4231-2

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, *405*(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

McCray, C. D., Atallah, E. H., & Gyakum, J. R. (2019). Long-duration freezing rain events over North America: Regional climatology and thermodynamic evolution. *Weather and Forecasting*, *34*(3), 665–681. https://doi.org/10.1175/WAF-D-18-0154.1

McCray, C. D., Gyakum, J. R., & Atallah, E. H. (2020). Regional thermodynamic characteristics distinguishing long- and short-duration freezing rain events over North America. *Weather and Forecasting*, *35*(2), 657–671. https://doi.org/10.1175/waf-d-19-0179.1

Mearns, L., McGinnis, S., Korytina, D., Arritt, R., Biner, S., Bukovsky, M., Chang, H.-I., Christensen, O., Herzmann, D., Jiao, Y., Kharin, S., Lazare, M., Nikulin, G., Qian, M., Scinocca, J., Winger, K., Castro, C., Frigon, A., & Gutowski, W. (2017). *The NA-CORDEX dataset, version 1.0*. NCAR Climate Data Gateway. https://doi.org/10.5065/D6SJ1JCH

Milton, J., & Bourque, A. (1999). *A climatological account of the January 1998 ice storm in Quebec* (Tech. Rep. No. CES-Q99-01). Environment Canada.

Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y., & Ludwig, R. (2019). Detecting climate change effects on Vb cyclones in a 50-member single-model ensemble using machine learning. *Geophysical Research Letters*, *46*(24), 14653–14661. https://doi.org/10.1029/2019GL084969

Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., Prabhat, M., & Pal, C. (2017). Extreme weather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Advances in Neural Information Processing Systems*, https://arXiv.org/abs/1612.02095

Ramer, J. (1993, 1–6 August). *An empirical technique for diagnosing precipitation type from model output*. Preprints of the fifth international conference on aviation weather systems (pp. 227–230), Vienna, VA, United States. American Meteorological Society.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ressler, G. M., Milrad, S. M., Atallah, E. H., & Gyakum, J. R. (2012). Synoptic-scale analysis of freezing rain events in Montreal, Quebec, Canada. *Weather and Forecasting*, *27*(2), 362–378. https://doi.org/10.1175/WAF-D-11-00071.1

Roebber, P. J., & Gyakum, J. R. (2003). Orographic influences on the mesoscale structure of the 1998 ice storm. *Monthly Weather Review*, *131*(1), 27–50. https://doi.org/10.1175/1520-0493(2003)131%3C0027:OIOTMS%3E2.0.CO;2

St-Pierre, M., Thériault, J. M., & Paquin, D. (2019). Influence of the model horizontal resolution on atmospheric conditions leading to freezing rain

in regional climate simulations. *Atmosphere-Ocean*, *57*(2), 101–119. https://doi.org/10.1080/07055900.2019.1583088

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., & Roeckner, E. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*(2), 146–172. https://doi.org/10.1002/jame.20015

Stuart, R. A., & Isaac, G. A. (1999). Freezing precipitation in Canada. *Atmosphere-Ocean*, *37*(1), 87–102. https://doi.org/10.1080/07055900.1999.9649622

Šeparović, L., Alexandru, A., Laprise, R., Martynov, A., Sushama, L., Winger, K., Tete, K., & Valin, M. (2013). Present climate and climate change over North America as simulated by the fifth-generation Canadian Regional Climate Model. *Climate Dynamics*, *41*(11), 3167–3201. https://doi.org/10.1007/s00382-013-1737-5

Valverde-Albacete, F. J., & Peláez-Moreno, C. (2014). 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox. *PloS one*, *9*(1), e84212–e84217. https://doi.org/10.1371/journal.pone.0084217

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Sénési, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Douville, H., Planton, S., Saint-Martin, D., Tyteca, S., Alkama, R., Belamari, S., Braun, A., & Chauvin, F. (2013). The CNRM-CM5.1 global climate model: Description and basic evaluation. *Climate Dynamics*, *40*(9-10), 2091–2121. https://doi.org/10.1007/s00382-011-1259-y

2.3 Paper III: Identifying the Atmospheric Drivers of Drought and Heat Using a Smoothed Deep Learning Approach

Reference: Mittermeier, M., Weigert, M., & Rügamer, D. (2021): Identifying the atmospheric drivers of drought and heat using a smoothed deep learning approach. *arXiv preprint*. arXiv:2111.05303.

Transition to paper III: Paper III is a workshop paper that was published as preprint and was accepted for the NeurIPS workshop: *Tackling Climate Change with Machine Learning*. It is a pre-study of paper IV. The study focuses on six selected circulation types of the subjective circulation type classification over Europe called *Großwetterlagen* by Hess & Brezowsky. These six types are related to the occurrence of heat and drought in Central Europe. A region that was only recently hit by combined heat and drought events in the years 2003, 2010 and 2018. The presented automatization of the established subjective Hess & Brezowsky classification using deep learning can help to study the influence of climate change on the driving circulation patterns of these hydro-meteorological extreme events. In contrast to paper I and II this study involves a multi-class classification. The workshop paper introduces modeling strategies to address challenges specific to the classification of circulation types, e.g., noisy data, imbalanced class distribution, and especially the defined dwell time of circulation types of at least three days.

Author's contribution: MM designed the concept of the study and carried out the data handling and calculations. MW implemented the transition smoothing step. DR supervised the process. All authors contributed to the manuscript.

Status: published

Journal: arXiv preprint; Workshop at NeurIPS 2021: *Tackling Climate Change with Machine Learning*

Identifying the atmospheric drivers of drought and heat using a smoothed deep learning approach

Magdalena Mittermeier
Department of Geography
LMU Munich
m.mittermeier@lmu.de

Maximilian Weigert
Statistical Consulting StaBLab
Department of Statistics
LMU Munich
maximilian.weigert@stat.uni-muenchen.de

David Rügamer
Department of Statistics
LMU Munich
david.ruegamer@stat.uni-muenchen.de

Abstract

Europe was hit by several, disastrous heat and drought events in recent summers. Besides thermodynamic influences, such hot and dry extremes are driven by certain atmospheric situations including anticyclonic conditions. Effects of climate change on atmospheric circulations are complex and many open research questions remain in this context, e.g., on future trends of anticyclonic conditions. Based on the combination of a catalog of labeled circulation patterns and spatial atmospheric variables, we propose a smoothed convolutional neural network classifier for six types of anticyclonic circulations that are associated with drought and heat. Our work can help to identify important drivers of hot and dry extremes in climate simulations, which allows to unveil the impact of climate change on these drivers. We address various challenges inherent to circulation pattern classification that are also present in other climate patterns, e.g., subjective labels and unambiguous transition periods.

1 Introduction

In recent summers such as those of 2003, 2010 and 2018, Europe has been subject to particularly outstanding summer drought and heat events, which caused large economic and societal damage including heat-related deaths [1, 2]. The frequency and intensity of hot and dry extremes has recently increased and is projected to further increase due to climate change and rising global mean temperatures [3, 4].

Drivers of hot and dry extremes There are two key processes leading to drought and heat events: thermodynamic and dynamic factors. Thermodynamic factors involve, e.g., evaporation and the feedback between soil moisture and air temperature. Dynamic factors on the other hand describe the atmospheric drivers of heat and drought, which are mainly anticyclonic conditions and blocking [3]. While anticyclonic conditions go along with various high-pressure systems, blocking describes a particular, persistent high-pressure situation that is associated with the displacement of westerly winds and their accompanying weather systems [3, 5]. These atmospheric drivers of hot and dry extremes are part of the large-scale atmospheric circulation in the mid-latitudes, which control the weather and climate over Europe [6, 7]. Changes in the atmospheric circulation are complex with opposing processes and thus many open research questions remain [5, 8].

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021.

Classification of circulation patterns Objectively classifying the circulation patterns that are associated with hot and dry extremes is an important step towards a better understanding of how climate change affects the atmospheric drivers of potentially disastrous extreme events. Previous studies, e.g., [9, 10] have shown that deep learning approaches are powerful tools for the detection of extreme weather in climate simulations. In this study, we use a subjective catalog of circulation type classifications over Europe by Hess & Brezowsky [11, 12]. Our goal is to learn the categorization of six circulation types with anticyclonic conditions over Europe, which are associated with dry and hot summer conditions in Central Europe [13]. The classification of circulation types comes with various challenges that need to be addressed with appropriate modeling strategies. Challenges include noisy labels due to subjective expert choices in ambiguous climate situations [14], an imbalanced class distribution of labels, undefined transition days between successive circulation patterns, and a fixed dwell time of a circulation pattern of at least three consecutive days by its definition [12].

Our contribution In this work we propose a novel modeling procedure to address existing challenges in classifying anticyclonic circulation patterns. Especially in times of large ensembles of climate simulations that consist of dozens of model runs and thousands of model years, our study can help to efficiently analyse large climate simulations and be another piece of the puzzle to better understand changes in the atmospheric drivers of drought and heat. We will make our model and the prepared data available upon publication.

2 Data

The Hess & Brezowsky catalog contains a subjective categorization of circulation patterns created by experts manually labelling air pressure patterns over Europe into 29 classes. In this way, daily air pressure constellations are retrospectively assigned to one of these classes. A circulation pattern is, by definition, required to last at least three days. The six circulation patterns associated with heat and drought are (abbreviations originate from German): *Zonal ridge across Central Europe (BM), Norwegian Sea-Iceland high, anticyclonic (HNA), North-easterly anticyclonic (NEA), Fennoscandian high, anticyclonic (HFA), Norwegian Sea-Fennoscandian high, anticyclonic (HNFA), and South-easterly anticyclonic (SEA)* [15, 13]. The mean air pressure patterns for the six circulation patterns of interest are given in Figure 1 for the variables *sea level pressure* and *geopotential height at 500 hPa* (average values at 5500 meters height). For the analysis of heat and drought, the remaining 23 circulation types are assigned to a residual class. The frequencies of the six anticyclonic patterns are between 8.5% (BM) and 1.4% (HNFA), whereas the residual class comprises about 80% of the days.

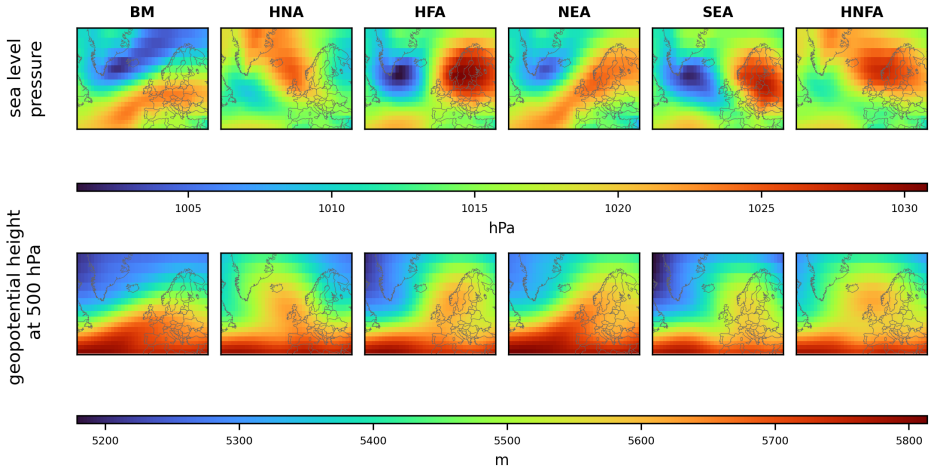

Figure 1: Mean air pressure patterns of the six anticyclonic circulation types BM, HNA, NEA, HFA, HNFA, and SEA (columns) averaged over all days in the period between 1900 and 2010. The plots are shown for the variables (rows) sea level pressure [hPa] and geopotential height at 500 hPa [m].

Next to the catalog by Hess & Brezowsky, we supplement our data base with the ERA-20C reanalysis data by the European Centre for Medium-Range Weather Forecasts [16]. The long record ERA-20C data set contains global spatial information on various climate parameters from 1900 to 2010. In accordance with the definition by [12], we use the two atmospheric variables, sea level pressure and geopotential height at 500 hPa as predictors for circulation patterns. The spatial domain of our data set is defined over a region covering Europe and the North Atlantic [12] (see Figure 1) with a spatial resolution of 5°x 5° resulting in a grid of 16 x 29 pixels. The resulting data set contains daily information over 111 years, i.e., T = 40541 observations.

3 Methods

Model definition, training, tuning and evaluation To account for the spatial information and the specific characteristics of circulation patterns, we use a convolutional neural network (CNN) following [17], who propose a network architecture for climate pattern detection problems. Since air pressure patterns of atmospheric features are comparatively simple, our chosen architecture consists of only two convolutional layers with larger kernels (5x5-8 and 5x5-16), a dropout layer and two fully-connected layers as well as individual channels for both climate parameters. The two different atmospheric variables are included as individual channels in the CNN as in [2, 9]. While there is reason to believe that accounting for the temporal structure of our data improves the model, a previous study [18] showed no improvement in the classification of circulation patterns when using a temporal-aware architecture. As explained in the next paragraph, our approach instead smoothes predicted labels to account for their temporal nature.

The model is trained using Adam with a batch size of 128 for 35 epochs and early stopping based on a validation set of size 3650 with patience of 6 epochs. Hyperparameter tuning for learning rate and dropout rate is performed using Bayesian optimization [19]. We evaluate the model using overall accuracy and macro F1-score. For class-specific evaluations, we consider recall and precision. To obtain performance estimates that are as unbiased as possible, a nested cross-validation with ten inner and eleven outer folds is used. In order to not leak intra-year information, observations within the same year are required to belong to the same fold.

Modeling challenges Our approach takes into account several data-specific characteristics for circulation pattern data. First, we employ a loss-weighting scheme to account for imbalanced classes by weighting the classes with their inverse frequencies. Moreover, the assigned categories in the Hess & Brezowsky catalog can be noisy, in particular for transition days between two subsequent circulation patterns. This is due to the continuous movement of pressure systems while circulation types are discrete by definition and in-between states do not fit in one or the other class. We solve this problem by using label smoothing [20] for the first and last day of each occurrence of a specific circulation pattern. Finally, our target variable must adhere to the aforementioned definition of a circulation pattern, implying a pattern to last at least three days. A transition-smoothing step ensures that this three day rule is respected. In this step, the final predicted class $\tilde{y}_t$ at time point $t = 2, ..., T-2$, is given by

$$\tilde{y}_t = \begin{cases} \hat{y}_{t-1} & \text{if } \hat{y}_{t-1} = \hat{y}_{t+1} \text{ (Neighborhood Consistency)}, \\ \hat{y}_{t-1} & \text{if } \hat{y}_t = \hat{y}_{t+1} \wedge \hat{y}_{t-1} = \hat{y}_{t+2} \text{ (2-days Consistency)}, \\ m(\hat{\boldsymbol{\pi}}_{t-1}, \hat{\boldsymbol{\pi}}_{t+1}) & \text{if } \hat{y}_t = \hat{y}_{t+1} \wedge \hat{y}_{t-1} \neq \hat{y}_{t+1} \text{ (Transition Membership)}, \\ m(\hat{\boldsymbol{\pi}}_{t-1}, \hat{\boldsymbol{\pi}}_{t+2}) & \text{else}, \end{cases}$$

where $\hat{\boldsymbol{\pi}}_t$ denotes the predicted probability vector at time t, $\hat{y}_t = \arg\max \hat{\boldsymbol{\pi}}_t$ the predicted class prior to the transition-smoothing step, and

$$m(\boldsymbol{\pi}_s, \boldsymbol{\pi}_t) = \arg\max\{\boldsymbol{\pi}_{u^*}\} \text{ with } u^* = \underset{u \in \{s,t\}}{\arg\max}\{\max(\boldsymbol{\pi}_u)\}.$$

This guarantees consistency with the required three day rule and systematically replaces isolated single or two day-type predictions.

4 Results

Taking into account the aforementioned subjectivity of the circulation pattern catalog and the noisy labels, the overall performance of the proposed model is satisfactory and our proposed smoothing

approaches consistently improve the model across all classes. Our smoothed convolutional neural network classifier achieves an macro F1-score of 39.2% and an overall accuracy of 60.7% averaged over the test sets during nested cross-validation. Table 1 shows the corresponding confusion matrix together with the precision and recall. The best performance in terms of recall is achieved for the circulation patterns HNA and BM, the lowest performance for NEA. In absolute numbers, misclassifications mainly occur for residual class observations. Due to the proposed error weighting technique, we obtain larger recall than precision values except for the residual class.

Table 1: Confusion matrix of our proposed approach, averaged over the test sets in the nested cross-validation. Correctly classified classes are highlighted in bold.

		LABELS								
		BM	HNA	HFA	NEA	SEA	HNFA	Residual	$\sum$	Precision
OUTPUTS	BM	**208.8**	4.0	7.7	6.5	1.7	43.3	477.4	711.2	0.29
	HNA	11.9	**75.8**	3.7	4.5	7.1	3.7	204.9	311.5	0.24
	HFA	22.9	3.5	**41.9**	14.6	3.6	1.5	138.6	226.6	0.19
	NEA	10.4	2.8	10.4	**61.5**	6.3	11.7	85.9	188.9	0.33
	SEA	3.2	11.6	3.6	15.0	**25.2**	5.5	78.4	142.5	0.18
	HNFA	9.1	3.8	2.4	20.5	5.5	**44.7**	185.7	271.7	0.17
	Residual	43.3	7.0	6.6	5.2	2.7	5.3	**1729.9**	1800	0.96
	$\sum$	309.6	108.6	76.2	127.7	52.1	77.4	2900.7	**3652.4**	–
	Recall	0.67	0.70	0.55	0.48	0.48	0.58	0.60	–	–

Table 2 shows an ablation study of our modeling procedure. Including the three day transition-smoothing step improves the overall accuracy by 4 percentage points and the macro F1-score by 3 percentage points. The class specific F1-scores also considerably increase for all patterns. A model without label- and transition-smoothing results in similar performance to the model without transition-smoothing, indicating that the proposed transition-smoothing is the key to our observed performance gains.

Table 2: Comparison of class-specific F1-scores (first 7 columns), accuracy and macro F1-Score (last two columns) for the final proposed model (Final), a model without transition-smoothing (No TS) and a model without label-smoothing (No LS and TS). Best results are highlighted in bold.

		BM	HNA	HFA	NEA	SEA	HNFA	Residual	**Accuracy**	**F1-score**
MODEL	Final	**0.41**	**0.36**	**0.28**	0.39	**0.26**	**0.26**	**0.74**	**0.60**	**0.38**
	No TS	0.39	0.33	0.25	0.36	0.23	0.23	0.70	0.56	0.36
	No LS and TS	0.39	0.33	0.25	**0.40**	0.26	0.23	0.69	0.56	0.36

5 Conclusion and Outlook

Our results indicate the high potential of deep learning-based methods in classifying the atmospheric drivers of drought and heat. We also demonstrate the effectiveness of our approaches to deal with typical challenges in circulation type classifications, e.g., label-smoothing for transition days and additional transition-smoothing for historical dwell time definitions. To the best of our knowledge, we are the first to use air pressure patterns over Europe to classify circulation patterns associated with drought and heat as given in the Hess & Brezowsky catalog. While the proposed approach can potentially also be used for other circulation patterns associated with different kinds of extreme climate events, our goal was to establish a baseline model for this specific and highly relevant circulation pattern categorization. Although our architecture follows a common standard in the field

of climatology, there is room for improvement in modeling the analyzed patterns. A possible future research direction is to evaluate further and more complex network structures, e.g., a ConvLSTM also taking into account the temporal dependence structure of the target. An overview of potential network architectures for spatio-temporal data settings is given by [21]. A more elaborate future approach is to investigate a deep hidden Markov model accounting for the state dwell times by assuming a latent process that emulates the three day rule inherent in this type of data.

Acknowledgements

We thank the anonymous reviewers for their constructive comments, which helped us to improve the manuscript.

The work of MM is funded by the Bavarian State Ministry for the Environment and Consumer Protection, the work of MW and DR by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A.

References

[1] A. Bastos, Z. Fu, P. Ciais, P. Friedlingstein, S. Sitch, J. Pongratz, U. Weber, M. Reichstein, P. Anthoni, A. Arneth, V. Haverd, A. Jain, E. Joetzjer, J. Knauer, S. Lienert, T. Loughran, P. C. McGuire, W. Obermeier, R. S. Padrón, H. Shi, H. Tian, N Viovy, and S. Zaehle. "Impacts of extreme summers on European ecosystems: a comparative analysis of 2003, 2010 and 2018". In: *Philosophical Transactions of the Royal Society B* 375.1810 (2020).

[2] B. Liu X.and He, L. Guo, L. Huang, and D. Chen. "Similarities and Differences in the Mechanisms Causing the European Summer Heatwaves in 2003, 2010, and 2018". In: *Earth ´s Future* 8.4 (2020), p. 5.

[3] L. Suarez-Gutierrez, W. A. Müller, C. Li, and J. Marotzke. "Dynamical and thermodynamical drivers of variability in European summer heat extremes". In: *Climate Dynamics* 54.9-10 (2020), 4351–4366.

[4] J. Spinoni et al. "Future Global Meteorological Drought Hot Spots: A Study Based on CORDEX Data". In: *Journal of Climate* 33.9 (2020), pp. 3635–3661.

[5] V. Masson-Delmotte, P. Zhai, S.L. Connors A. Pirani, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou. "Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change". In: *EGU General Assembly Conference Abstracts*. Cambridge University Press, 2021.

[6] M. F. Huguenin, E. M. Fischer, S. Kotlarski, S. C. Scherrer, C. Schwierz, and R. Knutti. "Lack of Change in the Projected Frequency and Persistence of Atmospheric Circulation Types Over Central Europe". In: *Geophysical Research Letters* 47.9 (2020).

[7] T. Woollings, A. Hannachi, and B. Hoskins. "Variability of the North Atlantic eddy-driven jet stream". In: *Journal of the Royal Meteorological Society* 136 (2010), pp. 856–868.

[8] M. Stendel, J. Francis, R. White, P. D. Williams, and T. Woollings. "The jet stream and climate change". In: *Climate Change* (2021), pp. 327–357.

[9] E. Racah, C. Beckham, T. Maharaj, S. E. Kahou, and C. Prabhat Pal. "ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events". In: (2016).

[10] T. Kurth, J. Zhang, N. Satish, I. Mitliagkas, E. Racah, M. A. Patwary, T. Malas, N. Sundaram, W. Bhimji, M. Smorkalov, J. Deslippe, M. Shiryaev, S. Sridharan, Prabhat, and P. Dubey. "Deep Learning at 15PF: Supervised and Semi-Supervised Classification for Scientific Data". In: (2017).

[11] P. Hess and H. Brezowsky. *Katalog der Großwetterlagen Europas (1881-2009)*. 1969.

[12] P. C. Werner and F. W. Gerstengarbe. *Katalog der Großwetterlagen Europas (1881-2009) nach Paul Hess und Helmut Brezowsky*. Nr. 119, 7. Aufl. PIK Report. Potsdam, 2010.

[13] H. Caspary. *Die Entwicklung von trockenen Großwetterlagen mit Auswirkungen auf den süddeutschen Raum*. Heft 18. Arbeitskreis KLIWA, 2012.

[14] R. Huth, C. Beck, A. Philipp, M. Demuzere, Z. Ustrnul, M. Cahynová, J. Kysel, and O. E. Tveitof. "Classifications of atmospheric circulation patterns: recent advances and applications". In: *Annals of the New York Academy of Sciences* 1146 (2008), pp. 105–152.

[15] P. Sýkorová and R. Huth. "The applicability of the Hess–Brezowsky synoptic classification to the description of climate elements in Europe". In: *Theoretical and Applied Climatology* 142.3-4 (2020), 1295–1309.

[16] P. Poli, H. Hersbach, D. P. Dee, P. Berrisford, A. J. Simmons, F. Vitart, P. Laloyaux, D. G. H. Tan, C. Peubey, J.-N. Thépaut, Y. Trémolet, E. V. Hólm, M. Bonavita, L. Isaksen, and M. Fisher. "ERA-20C: An atmospheric reanalysis of the twentieth century". In: *Journal of Climate* 29.11 (2016), pp. 4083–4097.

[17] Yunjie Liu, Evan Racah, Joaquin Correa, Amir Khosrowshahi, David Lavers, Kenneth Kunkel, Michael Wehner, William Collins, et al. "Application of deep convolutional neural networks for detecting extreme weather in climate datasets". In: *arXiv preprint arXiv:1605.01156* (2016).

[18] H. Funk, C. Becker, A. Hofheinz, G. Xi, Y. Zhang, F. Pfisterer, M. Weigert, and M. Mittermeier. "Towards an automated classification of Hess & Brezowsky's atmospheric circulation patterns Tief and Trog Mitteleuropa using Deep Learning Methods". In: *Environmental Informatics: A bogeyman or saviour to achieve the UN Sustainable Development Goals?* Ed. by V. Wohlgemuth, S. Naumann, H.K. Arndt, and G. Behrens. Shaker, forthcoming.

[19] J. Snoek, H. Larochelle, and R. P. Adams. "Practical bayesian optimization of machine learning algorithms". In: *Advances in neural information processing systems* 25 (2012).

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.

[21] S. Wang, J. Cao, and P. Yu. "Deep learning for spatio-temporal data mining: A survey". In: *IEEE transactions on knowledge and data engineering* (2020).

2.4 Paper IV: A Deep Learning based Classification of Atmospheric Circulation Types over Europe: Projection of Future Changes in a CMIP6 Large Ensemble

Reference: Mittermeier, M., Weigert, M., Rügamer, D., Küchenhoff, H. & Ludwig, R. (2022): A Deep Learning based Classification of Atmospheric Circulation Types over Europe: Projection of Future Changes in a CMIP6 Large Ensemble. *Environmental Research Letters*. submitted.

Transition to paper IV: Paper IV introduces an automated, deep-learning based classification of all 29 circulation types called *Großwetterlagen* over Europe from the subjective catalog by Hess & Brezowsky. These 29 circulation types are commonly used to study the dynamic drivers of various extreme events in Europe, e.g., heavy precipitation, floods and heat waves. The deep learning classifier involving a CNN architecture is designed for an application to large climate ensembles and surpasses existing automatizations of the 29 circulation types in classification accuracy. It is applied to the SMHI-LENS, a SMILE from the CMIP6 generation with 50 members, in order to study climate change effects on circulation type occurrence while considering the substantial uncertainty of internal climate variability. Surpassing the scopes of paper I and II, this paper addresses a multi-class classification task with a considerable number of classes. It furthermore deals with an uncertainty assessment of the deep learning method.

Author's contribution: The study concept was designed by MM. MM performed the data handling, calculations and analysis in close exchange with MW. MW implemented the smoothing step. DR supervised the deep learning procedure. HK and RL monitored and supported the entire research process. MM wrote the first manuscript version. All authors added valuable input and expertise to the final document.

Status: submitted

Journal: Environmental Research Letters (ERL)

Impact Factor: 6.79

A Deep Learning based Classification of Atmospheric Circulation Types over Europe: Projection of Future Changes in a CMIP6 Large Ensemble

Abstract. High- and low pressure systems of the large-scale atmospheric circulation in the mid-latitudes drive European weather and climate. Potential future changes in the occurrence of circulation types are highly relevant for society. Classifying the highly dynamic atmospheric circulation into discrete classes of circulation types helps to categorize the linkages between atmospheric forcing and surface conditions (e.g. extreme events). Previous studies have revealed a high internal variability of projected changes of circulation types. This requires the employment of a single-model initial-condition large ensemble and an automated classification method, which can be applied to large climate data sets. One of the most established classifications in Europe are the 29 subjective circulation types called *Großwetterlagen* by Hess & Brezowsky. We developed, in the first analysis of its kind, an automated version of this subjective classification using deep learning. Our classifier reaches an overall accuracy of 41.1 % on the test sets of nested cross-validation. It outperforms the state-of-the-art automatization of the HB circulation types in 22 of the 29 classes. We apply the deep learning classifier to the SMHI-LENS, a SMILE of the Coupled Model Intercomparison Project phase 6 (CMIP6), composed of 50 members of the EC-Earth3 model under the SSP37.0 scenario. For the analysis of future frequency changes of the 29 circulation patterns, we use the signal-to-noise ratio to discriminate the climate change signal from the noise of internal variability. Using a 5 %-significance level, we find significant frequency changes in 69 % of the circulation types when comparing the future (2071 to 2100) to a reference period (1991 to 2020).

Submitted to: *Environ. Res. Lett.*

1. Introduction

Large-scale atmospheric circulation in the mid-latitudes drives European weather and climate through the westerly jet stream and high- and low pressure systems originating from it (Huguenin et al., 2020; Woollings et al., 2010). Classifying the highly dynamic atmospheric circulation into discrete classes has been a key effort in synoptic climatology to gain a better understanding of the linkages between atmospheric forcing and surface conditions. Various different circulation type classifications exist. These can be categorized as subjective (manual), hybrid (mixed) or objective (automated/computer-assisted). Every classification consists of two steps: the class definition and the allocation

of pressure fields to these classes. For subjective classifications, the classes are manually defined by experts a priori to the assignment step, which is then also carried out manually. Hybrid methods are based on subjective class definitions with automized assignment steps. For objective methods, in contrast, the entire procedure is carried out in a numerical, automated way (Huth et al., 2008).

One of the most established classification schemes in Europe comprises 29 circulation types called *Großwetterlagen* by Hess & Brezowsky (HB circulation types; Hess and Brezowsky, 1952). Werner and Gerstengarbe (2010) published a revised catalog that covers the period from 1881-2009 and provides daily information on the HB circulation types. The catalog is constantly updated by the German Weather Service (DWD). Even though the subjectiveness of the HB circulation types involves considerable disadvantages in terms of inconsistencies and ambiguous class assignments, the main advantages of this classification are its intuitive naming convention and its high quality, e.g. for the description of climate elements especially over Central Europe (Sýkorová and Huth, 2020). The main benefits compared to an automated classification (e.g. through cluster analysis) are the abilities to describe real synoptic features and to also capture rare but relevant synoptic types (e.g. a specific type of blocking anticyclones; James, 2006b).

Due to these reasons the HB circulation types have been widely used for applications that study the connection between atmospheric circulation and extreme events (Sýkorová and Huth, 2020); this includes heavy rainfall (Minářová et al., 2017), floods (Petrow et al., 2009), extreme temperatures (Sulikowska and Wypych, 2020) and heat waves (Hoy et al., 2020). The impact of the HB circulation types on weather exposed sectors like renewable energies has also been investigated (Drücke et al., 2021). Sulikowska and Wypych (2020) discovered that most of the hot days of the exceptionally hot summer of 2019 in Europe occurred in connection with only four dominant HB circulation types. Petrow et al. (2009) identified a few circulation types that trigger the majority of flood events in Germany and found that some of these types significantly increased during the period from 1952 to 2002. Through analyzing historic trends, Hoffmann and Spekat (2021) found that wet- and dry HB circulation types have significantly changed in frequency and duration from 1961 to 2018, and suggest that changes in European rainfall patterns are largely caused by dynamical changes of circulation types.

Because of the connections between extreme events or climate variables of interest and driving circulation types, it is highly relevant to understand future changes in the occurrence of circulation types in the context of climate change. Huguenin et al. (2020) studied dynamic changes of large-scale atmospheric circulation types that are based on the HB circulation types and summarized them in ten groups of atmospheric flow (Beck et al., 2007). Using a multi-model ensemble, they found no clear future trend in frequency or persistence of the circulation types, and explained this with the large influence of both internal variability and model spread between different climate models (Huguenin et al., 2020). Due to its dynamic nature, the large-scale atmospheric circulation is highly variable. For the detection of future changes in circulation patterns, it is therefore essential to consider the range of internal variability of the climate system (Vautard

et al., 2016). While HB circulation types have been widely used in conjunction with historic data, only James (2006a) and Ringer et al. (2006) have examined future changes of all 29 HB circulation types in climate models. They use an automated (hybrid) version of the HB circulation types developed by James (2006b). This automated version uses climate mean composite plots (separately for winter and summer) of all 29 circulation types based on daily mean fields of sea level pressure (slp) and geopotential height at 500 hPa (z500). A specific day in the climate model is assigned to the HB circulation type whose composite field has the highest correlation coefficient to the smoothed mean pressure field of the given day. Using this method, James (2006a) found no clear trends for future circulation changes in HadGEM1 climate model runs and attributed this to the high interannual variability. James (2006a) states that a large database is needed in order to derive robust statements about changes in the European circulation patterns. In summary, this shows the relevance of the HB circulation types for extreme events and weather exposed sectors in Europe, but also the lack of knowledge regarding future changes of these circulation types due to the high uncertainty of internal variability.
In this paper, we introduce a new automated (hybrid) version of the classification of *Großwetterlagen* by Hess and Brezowsky (1952) using deep learning. The code of this classification method is published open-source (see Data availability statement) and enables the classification of large climate ensembles. The application to a single-model initial-condition large ensemble (SMILE) allows us to investigate changes in the occurrence of the 29 HB circulation types under climate change conditions while considering the highly relevant influence of internal variability. A SMILE contains several simulations (members) of one climate model that only differ in their initialization. Thus, the members are equally likely realizations of the future climate and span the uncertainty range of internal variability (Deser et al., 2012; Maher et al., 2021). Deep learning is the state of the art method for visual pattern recognition, which has been applied to different climate pattern classification and detection problems (Racah et al., 2016; Liu et al., 2016; Kurth et al., 2017; Mittermeier et al., 2019; Huntingford et al., 2019). It can help to utilize the expert knowledge contained in the long historic record of subjective classifications and provide an automated version that is appropriate for handling large data sets.

2. Data & Methods

2.1. Training data set

We train our deep learning classifier on historic examples of HB circulation types for the period from 1900 to 1980. The supervised training process is based on two data components. First, the catalog of *Grosswetterlagen* over Europe by Hess & Brezowsky (Werner and Gerstengarbe, 2010) contains a list of daily class affiliations for the 29 HB circulation types since 1900 derived from a manual classification of observed atmospheric pressure constellations. We use the catalog's class affiliations as labels for the training of

our deep neural network. Table 1 lists the 29 circulation patterns with their acronyms and definitions. The second data component is the ERA-20C reanalysis by the European Centre for Medium-Range Weather Forecasts (Poli et al., 2016) covering the period from 1900 to 2010. This data contains the spatial atmospheric pressure patterns that match to the labels from the catalog and are interpreted as images due to their pixelwise structure. We use the variables slp and z500 in a 5° spatial resolution over a domain covering Europe and parts of the North Atlantic (30-75°N, -65-45°O) based on Werner and Gerstengarbe (2010). Due to an implausible sudden discontinuity of the labels of the catalog that starts around the mid-1980s with an artificial increase in circulation type persistence (Kučerová et al., 2017), the time period from the year 1980 on is excluded and only the consistent data from 1900 to 1980 is used for training. The training database contains 29585 training examples of daily, historic HB circulation types. Figure 1 illustrates the typical air pressure constellations for each of the 29 classes for slp and z500.

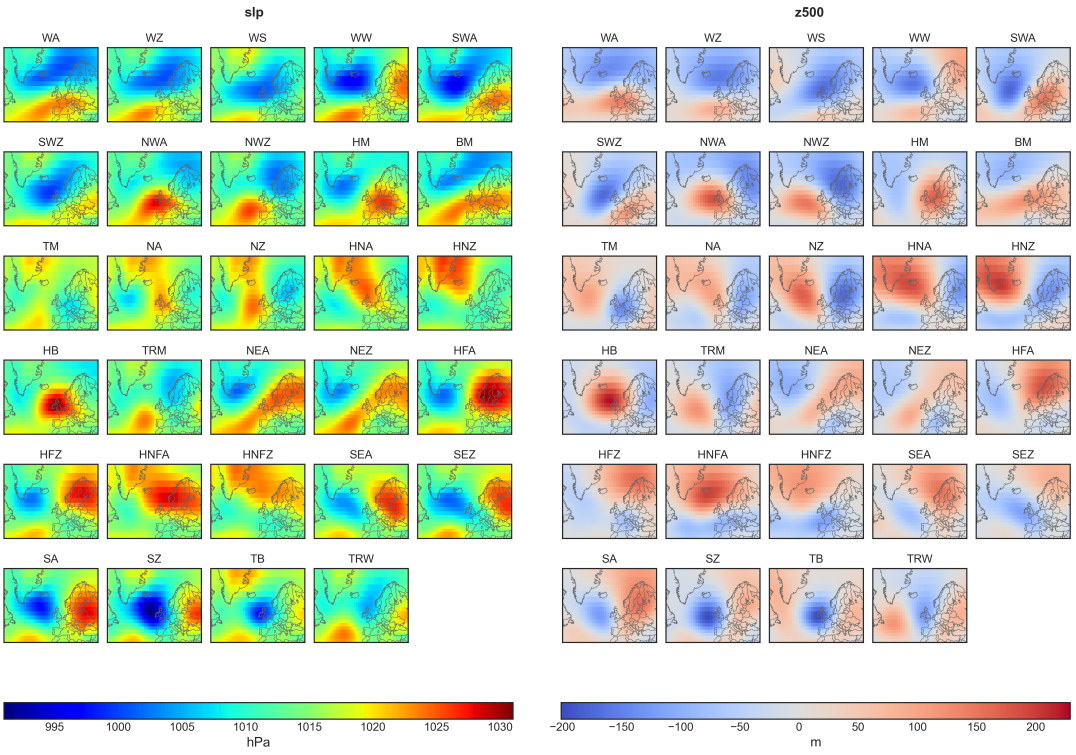

Figure 1: Typical air pressure constellations of the 29 circulation types averaged over all training examples (1900-1980) for sea level pressure (left) and geopotential height at 500 hPa (left). For slp we show the mean absolute pattern. In the case of z500, we show deviations from the mean, which give a more informative picture.

Table 1: List of the 29 circulation patterns with their acronyms, original German name and translated English name due to James (2006b).

Acronym	Original name (german)	Translated name (english)
WA	Westlage, antizyklonal	Anticyclonic Westerly
WZ	Westlage, zyklonal	Cyclonic Westerly
WS	Südliche Westlage	South-Shifted Westerly
WW	Winkelförmige Westlage	Maritime Westerly (Block E. Europe)
SWA	Südwestlage, antizyklonal	Anticyclonic North-Westerly
SWZ	Südwestlage, zyklonal	Cyclonic South-Westerly
NWA	Nordwestlage, antizyklonal	Anticyclonic North-Westerly
NWZ	Nordwestlage, zyklonal	Cyclonic North-Westerly
HM	Hoch Mitteleuropa	High over Central Europe
BM	Hochdruckbrücke (Rücken) Mitteleuropa	Zonal Ridge across Central Europe
TM	Tief Mitteleuropa	Low (Cut-Off) over Central Europe
NA	Nordlage, antizyklonal	Anticyclonic Northerly
NZ	Nordlage, zyklonal	Cyclonic Northerly
HNA	Hoch Nordmeer-Island, antizyklonal	Icelandic High, Ridge C. Europe
HNZ	Hoch Nordmeer-Island, zyklonal	Icelandic High, Trough C. Europe
HB	Hoch Britische Inseln	High over the British Isles
TRM	Trog Mitteleuropa	Trough over Central Europe
NEA	Nordostlage, antizyklonal	Anticyclonic North-Easterly
NEZ	Nordostlage, zyklonal	Cyclonic North-Easterly
HFA	Hoch Fennoskandien, antizyklonal	Scandinavian High, Ridge C. Europe
HFZ	Hoch Fennoskandien, zyklonal	Scandinavian High, Trough C. Europe
HNFA	Hoch Nordmeer-Fennoskandien, antizykl.	High Scandinavia-Iceland, Ridge C. Europe
HNFZ	Hoch Nordmeer-Fennoskandien, zyklonal	High Scandinavia-Iceland, Trough C. Europe
SEA	Südostlage, antizyklonal	Anticyclonic South-Easterly
SEZ	Südostlage, zyklonal	Cyclonic Southerly
SA	Südlage, antizyklonal	Anticylonic Southerly
SZ	Südlage, zyklonal	Cyclonic Southerly
TB	Tief Britische Inseln	Low over the British Isles
TRW	Trog Westeuropa	Trough over Western Europe

2.2. Network architecture and configuration

Our classification approach builds upon the image-like structure of the circulation patterns and uses a convolutional neural network (CNN). The architecture is an adaptation of the model provided by Liu et al. (2016) in the context of weather pattern detection and consists of two convolutional layers, a dropout layer and two-fully connected layers. In the convolutions, we use two individual channels for the climate parameters (slp and z500). Based on the original definition by Hess and Brezowsky (1952), the circulation types have to last for at least three days. That is why we apply transition smoothing as a post-processing step and smooth out class predictions that last for less than three days (details in the Appendix). The model is trained using Adam optimization (Kingma and Ba, 2014) with a batch size of 128, for 35 epochs and early stopping with a patience of six epochs. Hyperparameter tuning for learning rate and dropout rate is performed using Bayesian optimization (Snoek et al., 2012). The performance of the model is evaluated using the overall accuracy and the macro F1-Score, which takes the average of the class-specific F1-Scores and has a value range from 0 to 1 (Lewis et al., 1996). To obtain reliable and robust performance estimates, we apply nested cross-validation (Cawley and Talbot, 2010) with eight outer folds (test sets) and seven inner folds (model tuning). For each inner fold and its best hyperparameter set, we train five networks with different random weight initializations. The performance measures are then averaged over the eight outer test sets and five networks. To account for the time series nature of circulation patterns, training examples from the same year are required to be in the same cross-validation fold. The nested cross-validation is repeated for the final network applicable to new data. It is trained on the entire data set from 1900 to 1980 using the best hyperparameter configuration of all inner folds.

2.3. Uncertainty assessment

Due to their complexity, neural network training and their prediction are subject to uncertainty. In order to quantify our model's uncertainty, we use a deep ensemble (Lakshminarayanan et al., 2017) by generating a large number of 30 networks based on different random weight initializations. Using this approach, we can quantify the variance of predictions and generate more robust class affiliations by applying all 30 networks to the data and calculating a weighted average prediction. The importance weighting of each of the 30 predictions is proportional to the F1-score of each network. We apply this step to the ERA-40 reanalysis from September 1957 to August 2002 (Uppala et al., 2005) for evaluation purposes in comparison to the method by James (2006b), which is the state-of-the-art for an objective classification of the HB circulation patterns.

2.4. Climate ensemble: SMHI-LENS

The deep ensemble is furthermore applied to the climate ensemble SMHI-LENS (Wyser et al., 2021). SMHI-LENS is a SMILE of the Swedish Meteorological and Hydrological Institute, with the EC-Earth model (version 3.3.1) and 50 members. The SMHI-LENS follows the protocol of the Coupled Model Intercomparison Project phase 6 (CMIP6). We chose the SMHI-LENS due to its high number of members and the high performance of the EC-Earth3 model in reproducing daily sea-level pressure circulations types. Cannon (2020) compared 15 general circulation models with two reanalysis data sets. The EC-Earth3 was found to be one of the best performing CMIP6 models in terms of reproduction of frequency and persistence of circulation types under the consideration of internal variability, especially over Europe (Cannon, 2020). The SMHI-LENS is available for the period from 1970 to 2100 for four different scenarios in a 0.7° spatial resolution. It uses the macro initialization method for the generation of its ensemble members. We use the strong climate scenario SSP37.0 and a daily resolution on the 5° grid over the Europe-North-Atlantic domain (see Chapter 2.1). Frequencies of occurrence of circulation patterns are compared for two 30-year periods, a far future horizon from 2071 to 2100, and the reference period from 1991 to 2020. The signal-to-noise ratio (S/N-ratio) and its significance is calculated based on Aalbers et al. (2018) using a two-sided t-test. The S/N-ratio states, if the forced response (ensemble averaged frequency change) exceeds the noise (standard deviation of the ensemble). To globally control the significance level of 5 % over all 29 circulation types, the method by Benjamini and Hochberg (1995) is used to adjust for multiple testing based on the false discovery rate.

3. Results

3.1. Method Evaluation

To evaluate the performance of our method, the daily class affiliations from the original HB circulation type catalog (Werner and Gerstengarbe, 2010) are compared to the class predictions of the deep learning classifier. On the outer folds of the nested cross-validation, we obtain a macro F1-score of 39.3 and an overall accuracy of 41.1 %. The class-specific F1-scores are given in the first column of Table 2. Table 2 further shows the performance measures when applying our deep learning classifier to the ERA-40 reanalysis as analyzed by James (2006b). The deep learning classifier outperforms the method by James (2006b) in 22 of the 29 classes. For the circulation patterns HM, NA and SZ, the performance is more than 10 % higher, while the approach by James (2006b) works especially well for NWA, TRM or HNFA. The overall accuracy of the deep learning method on ERA-40 is 43.04 % and 39.1 % for James (2006b). However, these results can be overly optimistic as the ERA-40 reanalysis data set has 24 years overlap with the training data set. A conservative performance measure of our method is given by the nested cross-validation results, while ERA-40 is only used to compare with James

(2006b). The confusion matrix showing the average classifications on the test sets during cross-validation is shown in Table A1 in the Appendix. Most of the misclassifications occur between pairs of anticyclonic- and cyclonic circulation.
The deep learning classifications are compared with the HB circulation type catalog in respect of the frequency distribution of the classes (see Appendix, Figure A1a). The network reproduces the relative order of the classes well, but clearly underestimates the class WZ. Also the classes HM and BM are underestimated by the network, while it overestimates the class WS. The climate ensemble SMHI-LENS reproduces the circulation types well (see Appendix, Figure A1b). Except for BM, for which the climate model overestimates the frequency, all boxplots cover or intersect with the frequencies in ERA-20C reanalysis data.

Figure 2 evaluates the "synoptic performance" of our deep learning classifier for each class (Verdecchia et al., 1996). The signature plots are derived by taking the average field of a certain class and subtracting the average field of all other situations from it. This shows, which synoptic characteristics distinguish a single class from the other classes. Signature plots are given for four different cases: labels, predictions, false positives and false negatives. When columns 1 and 2 are very similar, the average signature of the deep learning prediction agrees well with the average signature of the labels as derived from the original HB catalog. A further noteworthy insight is given when there is large agreement between column 1 and 3, which can indicate a misclassification in the catalog. In this case, the deep learning method might correctly classify the situation, while the labels disagree. Apart from slight differences for some classes, column 3 agrees generally very well with column 1. Exceptions for which the CNN does not correctly identify the patterns are for example: SEZ, for which the extent of the low pressure ridge in the North is too large; TRW, because its high pressure system is too pronounced and TB for which the extent of the low pressure system is too small. In other cases (marked by green class names), the signature plots are an indication to question the choice in the subjective catalog. Although the deep learning classifier only considers the information provided in the data, and subjective reasons for the label classification are not available, a certain level of arbitrariness in the catalog has already been recognized before by James (2006b). Column 4 reveals the false negatives of the deep learning classifier. In some cases (e.g. WA, WS, WW, NWZ and SEZ), the false negatives clearly differ from the signature pattern of the labels in column 1. In these cases, the catalog labels may be questioned and predictions of our approach seem plausible – at least based on the objective slp and z500 data.

Figure 3 depicts the uncertainty obtained through the deep ensemble (30 members) compared to the internal climate variability (50 members) as percentage of the total uncertainty. The network's uncertainty range lies at 11-33 % for the entire year. It is larger in winter and smaller in summer. Note that for the deep learning part this does not take the variability of hyperparameter tuning into account. In respect to typical

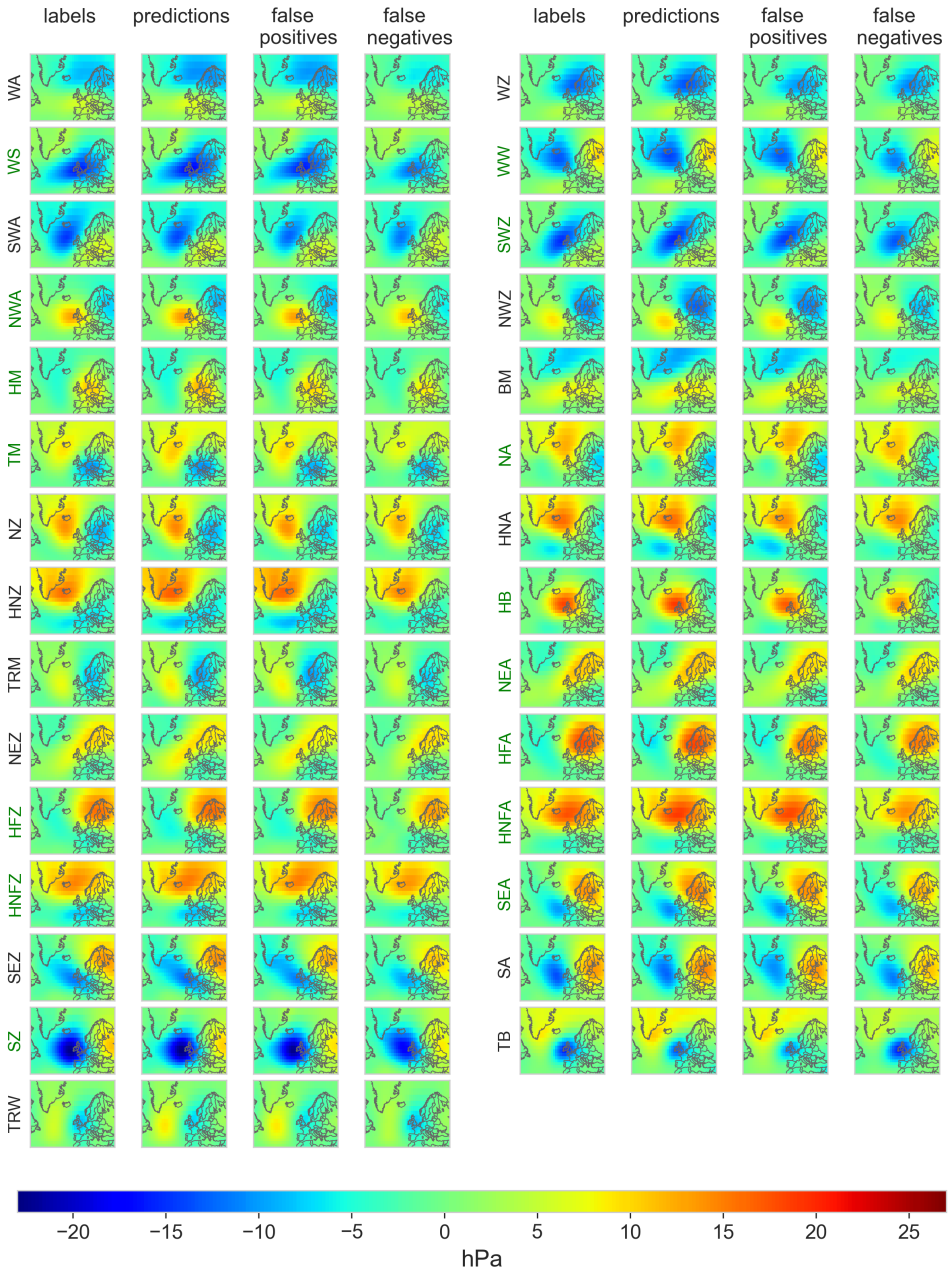

Figure 2: Signature plots of the 29 circulation patterns at slp. Each circulation type (CT) is shown in the respective row for four cases: column 1) labels: labels showing the indicated CT, column 2) predictions: deep ensemble predictions showing the indicated CT, column 3) false positives: signature pattern, when the deep ensemble predicts the indicated CT while labels state differently, column 4) false negatives: labels stating the indicated CT while deep ensemble predicts differently. The signature plots are derived by calculating the average of all days for which the conditions for this CT are met and subtracting the average of all other CTs. Thus, the composite plots show patterns that distinguish a certain CT from the other types. The green class labels indicate CT for which there is a good agreement between column 1 and 3.

climate modeling uncertainties (Hawkins and Sutton, 2011), the influence of climate model choice and scenario choice is not considered.

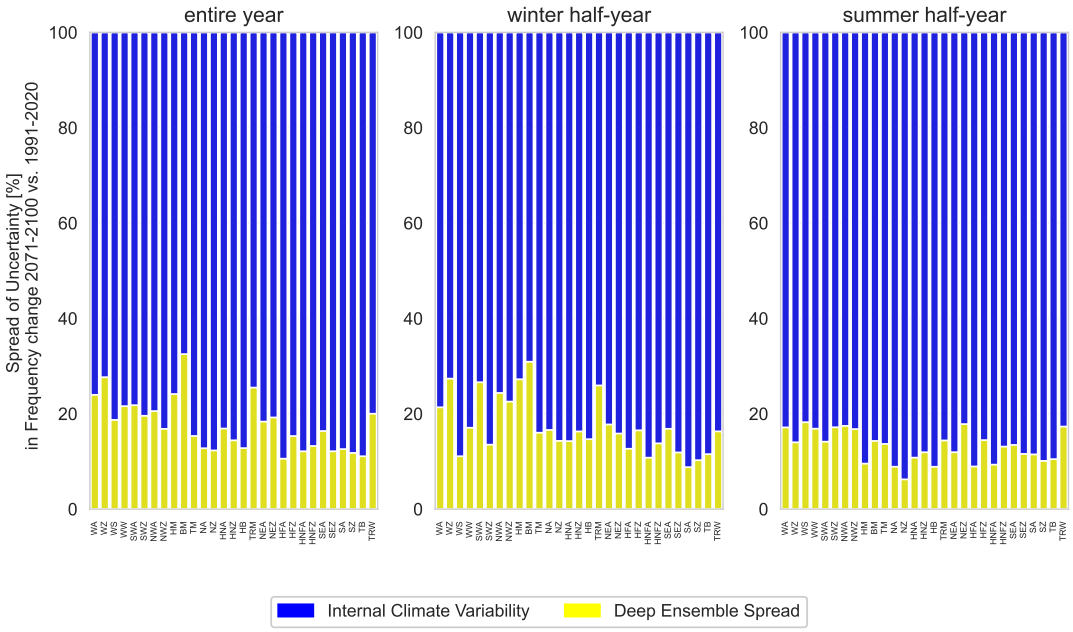

Figure 3: Stacked barplots illustrating the contribution of the two sources of uncertainty: deep model initialization and internal climate variability in [%] of the entire spread. The uncertainty sources are investigated for the projected frequency changes of the 29 circulation types between the far future (2071-2100) and the reference period (1991-2020) for the entire year, the winter half-year (ONDJFM) and the summer half-year (AMJJAS). The entire variance is based on 1500 values (50 climate models times 30 deep model predictions).

Table 2: Comparison of class-specific F1-scores of our Deep Learning classifier (DL) evaluated on nested cross-validation (CV), on ERA-40 reanalysis data and comparison to the classification method of James (2006b) on ERA-40. The best result on ERA-40 is highlighted in bold in order to facilitate the comparison of the two methods. The overall accuracy and macro F1-Scores are given in the last two rows.

Circulation type	F1-score DL CV	F1-score DL ERA-40	F1-score James
WA	44.6	**49.31**	40.04
WZ	47.08	**54.73**	52.5
WS	45.39	**42.27**	34.89
WW	37.7	**39.66**	29.91
SWA	35.36	**44.14**	36.44
SWZ	30.86	**41.24**	39.44
NWA	38.88	26.0	**33.51**
NWZ	37.07	41.1	**43.28**
HM	51.24	**56.74**	43.07
BM	47.29	**40.21**	37.88
TM	37.23	**37.4**	36.96
NA	24.85	**30.77**	15.82
NZ	44.32	**43.56**	41.31
HNA	45.57	**45.87**	45.55
HNZ	27.11	33.02	**36.53**
HB	50.99	**47.65**	44.78
TRM	27.86	31.32	**39.35**
NEA	41.44	**33.29**	29.74
NEZ	33.12	**33.22**	27.12
HFA	45.32	**46.89**	40.94
HFZ	24.81	31.56	**32.85**
HNFA	33.35	35.89	**43.21**
HNFZ	34.02	**39.38**	33.06
SEA	38.09	**32.2**	27.25
SEZ	37.93	**38.81**	31.01
SA	39.84	**43.73**	33.89
SZ	38.19	**41.31**	26.57
TB	42.11	**46.67**	37.7
TRW	29.34	35.57	**37.64**
Macro F1-score	38.3	**40.12**	36.28
Overall Accuracy	41.1	**43.04**	39.1

3.2. Future changes

We apply the weighted deep ensemble to the SMHI-LENS with its 50 members to quantify the spread of internal variability for absolute future frequency changes of the 29 circulation patterns between 2071 to 2100 and the reference period (1991-2020). Figure 4 shows the spread of internal variability for all circulation types for the entire year as well as the winter and summer half-year illustrated by boxplots. Significant changes in terms of the S/N-ratio are indicated with bold class names. Tables with the complete S/N-ratio values are given in the Appendix (Table A2 and Table A3). For most circulation types, the boxplots intersect with the horizontal line at zero and the members disagree in the sign of the trend. Overall, absolute changes are small and lie within a range of ± 5 days for most circulation types. This finding is in line with Huguenin et al. (2020), who find small changes of ± 4 days per season in a multi-model ensemble for 10 groups of circulation patterns that are based on the 29 HB circulation types. Note that for the circulation types TM-TRW (in the presented order), which have small absolute frequencies, changes of ± 5 days can still mean high relative changes of around ±50 % (see Apendix, Figure A2). For some circulation types, single members outside the interquartile range show relative changes of >50 %, especially in the winter half-year. Different to the studies by James (2006a) and Ringer et al. (2006), who analyzed all 29 circulation types in the climate model HadGEM1 and have not found significant frequency changes, our analysis of a SMILE allows to identify significant frequency changes due to climate change despite the high spread of internal variability. In Figure 4 d-f, we plot the class-specific F1-scores of our deep learning classifier and their range throughout the deep ensemble. This allows to take into account the quality of predictions for each class. Reliable statements can be made for the circulation patterns WA, WZ, WS, HM, HNA, HB, HFA, SA, and TB throughout the entire year. The clearest absolute climate trend is found for the anticyclonic westerly circulation (WA), which shows an increasing trend for the entire year (and the summer half-year) with a median of 6.6 days per year (summer: 5.4 days per year) and a S/N-ratio of 1.5 (summer: 1.7). For WA, the climate change signal clearly exceeds the noise of internal variability. The increasing winter trends of HFA and TB are also significant, as well as decreasing summer trends of WS, HB and SA and the increasing summer trend for HM.

In general, we find a decreasing trend for south-easterly circulations (SEA and SEZ) in both summer and winter (trends are significant except for SEA in winter), although their reliability based on F1-scores fluctuates seasonally. For winter, this goes in line with the findings by Herrera-Lormendez et al. (2021), who have detected a decreasing trend for south-easterly circulations from the Jenkinson-Collison classification using four members of EC-Earth3 under SSP58.5. The classification by Jenkinson and Collison (1977) is an automated version of the subjective Lamb catalog developed for the British Isles. Herrera-Lormendez et al. (2021) applied this classification to Europe and distinguished 11 circulation types. Our results also support the findings of Herrera-Lormendez et al. (2021) for the increasing summer trend of north-easterly circulations (in our case significant for

NEA) and the decreasing summer trend for Northerlies (in our case significant for NZ and HB).

Our results make clear that the spread of internal variability is tremendous and it is difficult to derive systematic changes of circulation patterns grouped by their wind directions. Despite the high internal variability, the results of the S/N-ratio are very clear, showing a significant change in 69 % of the classes for the total year, 34 % for the winter and 69 % for the summer half-year.

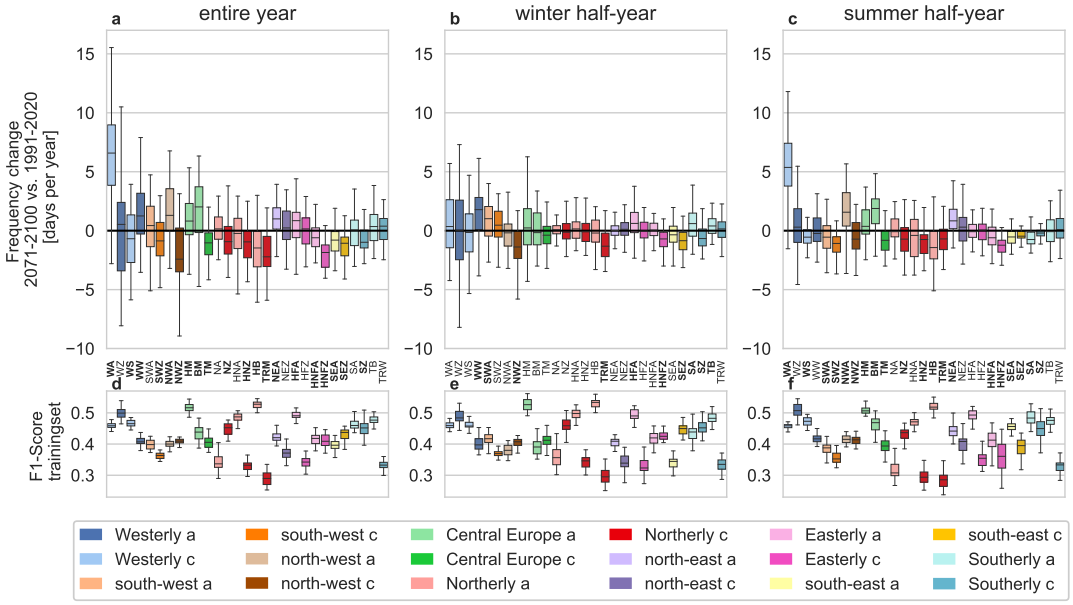

Figure 4: Upper plots a-c: Boxplots showing the absolute change in the frequency of occurrence [days per year] of the 29 circulation patterns between the far future 2071-2100 under the SSP37.0 scenario and the reference period 1991-2020 for the entire year, the winter half-year (ONDJFM) and the summer half-year (AMJJAS). The boxplots cover the distribution of the 50 members of the climate model ensemble SMHI-LENS. Bold class name on the x-axis indicate a significant S/N-ratio based on multiple testing with a significance level of 5 %. Lower plots d-f: Boxplots illustrating the spread of F1-scores of the 30 models of the deep ensemble trained on the entire training data set. Outliers outside of the boxplots' whiskers are not shown. The colors indicate groups of wind directions (Hoy et al., 2013). Pastel colors in the legend indicated by an "a" stand for anticyclonic, dark colors indicated with a "c" for cyclonic circulation.

4. Conclusion

In this study, we introduced a new automated classification method for the 29 circulation types defined by Hess and Brezowsky (1952) using deep learning. Our method shows the potential of deep learning in circulation type classification and outperforms the state-of-the-art method by James (2006b) in 22 of the 29 classes. We applied the deep learning classifier to a SMILE of the CMIP6 generation, the SMHI-LENS, which comprises 50 members of the EC-Earth3 general circulation model. Our study is the first one that analyzes future frequency changes of all 29 circulation patterns in a SMILE. Thus, other than previous studies on climate change impacts on the HB circulation types (James, 2006b; Ringer et al., 2006), we can identify significant frequency changes despite the high range of internal variability.

A better understanding of climate change impacts on the European circulation patterns is of high societal relevance due to their direct influence on our daily weather and the strong relation to extreme events like heavy rainfall (Minářová et al., 2017), floods (Petrow et al., 2009), hot days (Sulikowska and Wypych, 2020) and heat waves (Hoy et al., 2020). Our results show an immense spread of internal variability when investigating future frequency changes of the circulation patterns in the SMHI-LENS under the SSP37.0 scenario. Despite the high spread of internal variability, our results of the S/N-ratio show significant ($alpha = 5$ %) absolute frequency changes for a high number of classes (69 % of the classes for the entire year, 34 % for the winter half-year and 69 % for the summer half-year). This underlines the great benefit in using a SMILE when analyzing climate change effects on the highly dynamic large-scale atmospheric circulation over Europe. In absolute numbers, the frequency changes lie in a range of ± 5 days for most circulation types, which is in line with the findings by Huguenin et al. (2020). For the circulation types TM-TRW (in the presented order), which occur only on a few days per year, small absolute changes can still mean high relative changes (for some circulation types around ±50 %, for some members even >50 %). The most distinct absolute change is found for Anticyclonic Westerlies (WA) with an increasing trend for the entire year with a median of 6.6 days per year and a S/N-ratio of 1.5. Here, the climate change signal clearly exceeds the noise of internal variability.

The classification results show that our deep learning classifier can yield good predictions at low computational costs. This makes our method advantageous for application to large climate data sets such as multi-model ensembles or SMILEs. Towards the goal of reproducing the original subjective HB circulation types, it achieves higher performance measures as the method by James (2006b). For some classes, a larger part of the misclassifications of our deep learning classifier seem to be synoptically correct. The labels from the HB circulation type catalog (Werner and Gerstengarbe, 2010) are subjective and hold inconsistencies and ambiguous class affiliations (James, 2006b; Kučerová et al., 2017). This means that the labels taken as ground truth hold a certain, unquantified human level error. Our findings suggest that this human level error might be substantial for some classes.

The deep learning classifier is designed for the application to climate models, as this requires an automated version of the HB circulation type classification. It is not meant to replace a subjective continuation of the HB catalog and it is not suitable for this as long as the human level error is unquantified and there is potential to improve the performance of the classifier. A disadvantage of the deep learning approach is its potentially high variability, which can be caused by model uncertainty or too noisy data. Our evaluation shows that the variability of the deep learning method contributes up to 32.5 % of the entire variance when applying our method to the SMHI-LENS. To deal with this uncertainty, we use a deep ensemble of 30 networks with different initializations and calculate a performance-weighted mean of this deep ensemble when applying the classifier on new data.
Besides quantifying the human level error in the labels, possible future research could evaluate further network architectures for an improvement of the deep learning performance. Considering the temporal development of circulation patterns by using a temporal-aware ConvLSTM architecture might improve the classification accuracy. Furthermore, a deep hidden Markov model could improve the performance by including the three-day-definition of HB circulation types directly in the training process. In order to evaluate the uncertainties in frequency changes coming from different climate models and forcing scenarios, a combination of multi-model as well as single-model ensembles under different forcing scenarios is desirable. The deep learning classifier introduced in this study can serve as valuable tool for the analysis of such a comprehensive data set.

Data availability statement

The code for the application of our deep learning classifier will be published upon publication via zenodo (with a doi) and a GitHub-repository. The code is based on python (version 3.6). The ERA-40 and ERA-20C reanalysis are derived from the European Centre for Medium-Range Weather Forecasts (ECMWF): http://www.ecmwf.int/. The SMHI-LENS is publicly available from the data portal of the Earth System Grid Federation (ESGF): https://esgf-data.dkrz.de/.

Acknowledgments

The provision of the digital Hess & Brezowsky catalog for the years 1900-2010 by the German Weather Service is highly appreciated.

Appendix

Transition smoothing Our classifications must adhere to the definition that a circulation type lasts for at least three days. A transition-smoothing step ensures that this rule is respected by post-processing the time series of the network classifications. First, circulation types that last for less than three days are identified. Next, these transitions are tested for *neighborhood consistency* and *transition membership*. Neighborhood consistency describes the situation when before and after the transition the same circulation type occurs. The transition days are then smoothed by assigning this type to it. In case of transition membership different circulation types occur before and after the transition. Here, the transition days obtain the class affiliation of the circulation type before or after, depending on which class has the higher predicted probability.

Table A1: Confusion matrix of our proposed smoothed approach, averaged over the test sets in the nested cross-validation. Correctly classified classes are highlighted in bold.

		LABELS																												
		WA	WZ	WS	WW	SWA	SWZ	NWA	NWZ	HM	BM	TM	NA	NZ	HNA	HNZ	HB	TRM	NEA	NEZ	HFA	HFZ	HNFA	HNFZ	SEA	SEZ	SA	SZ	TB	TRW	∑	Precision
OUTPUTS	WA	**102**	62	1	3	4	2	10	8	22	26	0	1	1	0	0	1	3	1	1	0	0	0	0	0	0	0	0	1	3	253	0,40
	WZ	13	**195**	12	5	2	6	1	11	4	2	0	1	2	1	0	0	6	1	0	0	0	0	0	0	0	0	0	2	5	272	0,72
	WS	1	51	**67**	3	0	8	0	4	1	1	4	0	0	1	3	0	6	0	0	0	0	0	1	0	3	0	1	6	5	170	0,40
	WW	4	24	3	**42**	2	4	1	3	6	6	1	0	0	0	0	0	3	2	3	2	1	0	0	2	2	3	2	2	6	125	0,33
	SWA	14	18	1	4	**34**	11	0	0	25	7	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	4	0	1	2	125	0,27
	SWZ	4	36	9	5	8	**30**	0	1	4	1	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	2	1	4	5	115	0,26
	NWA	14	8	0	1	0	0	**58**	18	12	20	0	1	4	2	1	12	3	2	2	0	0	0	0	0	0	0	0	0	1	159	0,37
	NWZ	10	56	2	2	0	0	15	**67**	2	4	2	1	7	1	0	1	21	1	1	0	1	0	0	0	1	0	0	1	3	198	0,34
	HM	8	4	1	1	4	1	5	1	**142**	23	0	1	0	2	0	4	0	2	1	7	1	1	0	2	0	3	0	0	1	216	0,66
	BM	15	5	0	3	1	0	7	2	25	**104**	0	0	1	1	0	2	3	3	3	3	0	1	0	1	0	2	0	0	2	187	0,56
	TM	0	8	4	1	0	0	0	3	0	1	**34**	1	6	0	4	0	10	1	3	1	1	1	5	0	1	0	1	4	6	95	0,36
	NA	2	4	0	0	0	0	3	1	4	1	0	**12**	7	12	3	3	1	2	2	1	0	1	1	0	0	0	0	0	0	59	0,20
	NZ	2	7	0	0	0	0	6	12	1	2	6	4	**52**	4	5	3	15	1	2	0	0	0	1	0	0	0	0	0	1	125	0,42
	HNA	1	4	1	0	0	0	2	0	11	3	1	6	3	**54**	5	10	1	1	1	2	0	3	2	2	0	0	0	1	0	116	0,47
	HNZ	0	5	3	0	0	1	1	0	1	1	7	2	7	10	**17**	1	3	0	1	0	0	1	7	1	0	0	0	1	2	72	0,23
	HB	1	1	0	0	0	0	19	4	13	11	0	2	2	8	0	**65**	2	5	2	2	0	1	0	0	0	0	0	0	0	141	0,46
	TRM	3	14	2	1	0	1	1	17	1	4	6	0	9	1	0	0	**35**	1	2	0	0	0	0	0	1	0	0	1	8	109	0,32
	NEA	1	2	0	2	0	0	4	1	10	10	1	1	1	3	0	5	1	**43**	13	10	2	2	1	1	0	1	0	0	1	114	0,38
	NEZ	1	2	1	1	0	0	5	1	1	5	4	0	2	1	1	2	5	10	**26**	2	2	1	2	0	1	0	0	0	1	79	0,33
	HFA	1	1	0	1	1	0	0	0	21	5	0	0	0	3	0	1	1	9	2	**63**	5	4	1	9	2	4	0	0	1	135	0,47
	HFZ	0	1	1	2	0	0	0	1	1	1	3	0	0	0	0	0	1	3	8	12	**12**	1	4	2	5	1	0	1	1	61	0,20
	HNFA	0	1	0	0	0	0	0	0	2	3	1	1	2	11	1	2	1	3	3	12	1	**21**	8	3	1	0	0	1	1	79	0,27
	HNFZ	0	3	1	0	0	0	0	0	1	1	7	0	2	4	7	0	1	0	2	4	2	6	**24**	3	2	0	0	1	2	74	0,32
	SEA	0	1	1	1	1	0	0	0	7	2	0	0	0	3	0	0	0	0	1	12	2	2	3	**32**	9	8	1	1	1	87	0,36
	SEZ	0	2	3	4	0	0	0	1	0	1	2	0	0	0	0	0	2	1	2	1	4	0	3	6	**24**	2	3	2	3	68	0,35
	SA	1	2	0	3	4	3	0	0	16	5	0	0	0	1	0	0	0	0	0	6	0	0	0	7	1	**34**	4	1	3	94	0,36
	SZ	0	1	5	4	1	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4	8	**19**	7	5	63	0,30
	TB	2	20	7	4	2	5	0	3	3	1	3	0	0	0	1	0	3	0	0	1	0	0	1	1	2	2	2	**44**	12	120	0,37
	TRW	4	18	3	3	1	3	1	4	2	4	2	0	1	0	0	0	12	0	1	0	1	0	0	1	1	1	1	7	**32**	104	0,31
	∑	205	557	126	97	68	81	140	164	340	255	88	35	111	123	51	115	140	94	79	143	36	49	67	78	59	76	36	89	116	3616	–
	Recall	0,50	0,35	0,53	0,43	0,50	0,37	0,42	0,41	0,42	0,41	0,39	0,34	0,47	0,44	0,33	0,57	0,25	0,46	0,33	0,44	0,33	0,44	0,36	0,40	0,41	0,45	0,53	0,49	0,28	0,41	–

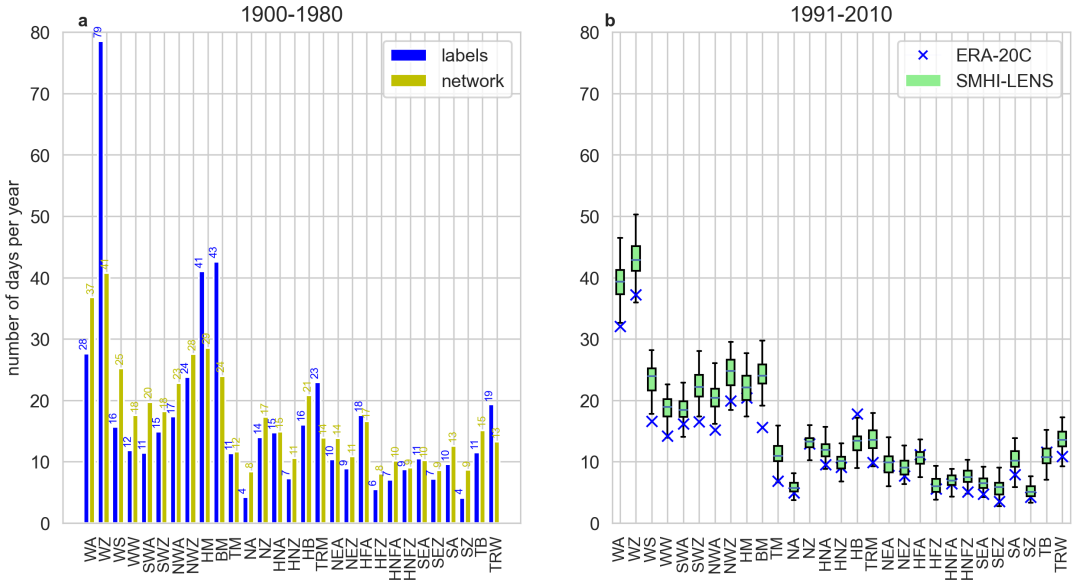

Figure A1: Frequency distribution of the 29 circulation types in number of days per year. Left: for the training period 1900-1980 in the HB circulation type catalog (labels) and in predictions of our deep learning classifier on ERA-20C reanalysis (network). Right: predictions of the deep learning classifier for the period 1991-2020 in ERA-20C reanalysis (blue) and in the 50 members of the SMHI-LENS (boxplots).

Table A2: S/N-ratios for the 15 circulation types WA-HNZ for the total year, winter half-year and summer half-year.

S/N-ratio	WA	WZ	WS	WW	SWA	SWZ	NWA	NWZ	HM	BM	TM	NA	NZ	HNA	HNZ
total	1.51	0.05	-0.31	0.53	0.15	-0.36	0.6	-0.7	0.31	0.76	-0.65	0.09	-0.36	-0.14	-0.65
winter half-year	0.23	-0.01	-0.11	0.69	0.53	0.16	-0.11	-0.68	0.13	0.05	-0.33	0.16	-0.03	0.19	-0.11
summer half-year	1.71	0.09	-0.57	-0.02	-0.39	-1.22	0.8	-0.43	0.34	0.86	-0.67	0.03	-0.39	-0.28	-0.73

Table A3: S/N-ratios for the 14 circulation types HB-TRW for the total year, winter half-year and summer half-year.

S/N-ratio	HB	TRM	NEA	NEZ	HFA	HFZ	HNFA	HNFZ	SEA	SEZ	SA	SZ	TB	TRW
total	-0.71	-0.93	0.68	0.23	0.34	0.01	-0.37	-1.62	-0.67	-1.09	-0.02	-0.66	0.29	0.24
winter half-year	-0.18	-0.85	0.08	0.22	0.4	0.05	0.06	-0.83	0.29	-0.81	0.42	-0.47	0.55	0.11
summer half-year	-0.73	-0.5	0.78	0.14	0.04	-0.04	-0.46	-1.71	-0.77	-0.9	-0.78	-0.6	-0.02	0.2

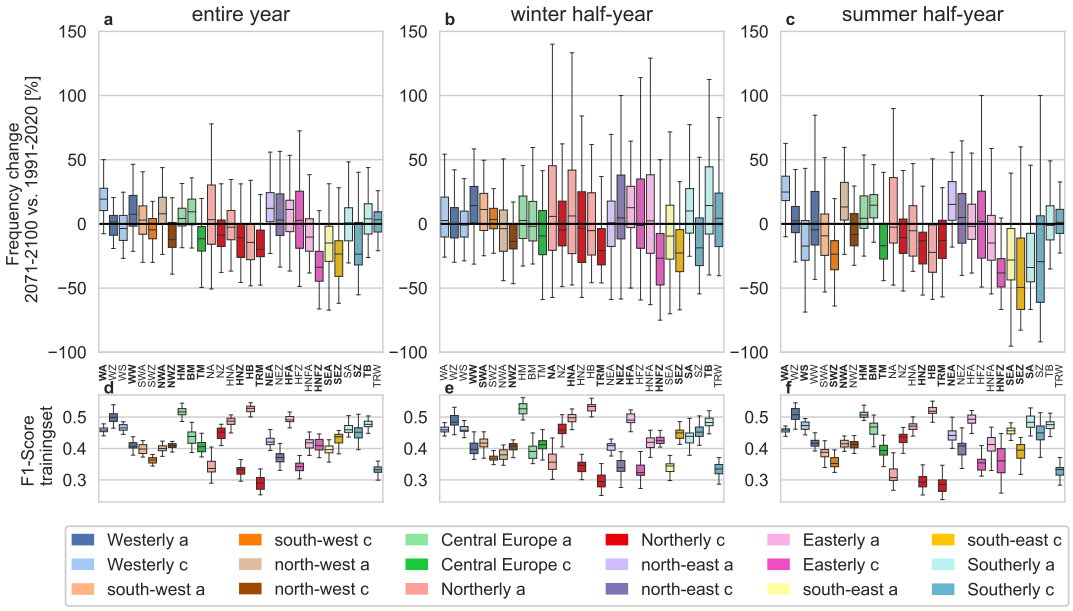

Figure A2: Upper plots a-c: Boxplots showing the change in the relative frequency of occurrence [%] of the 29 circulation patterns between the far future 2071-2100 under the SSP37.0 scenario and the reference period 1990-2020 for the entire year, the winter half-year (ONDJFM) and the summer half-year (AMJJAS). The boxplots cover the distribution of the 50 members of the climate model ensemble SMHI-LENS. Lower plots d-f: Boxplots illustrating the spread of F1-scores of the 30 models of the deep ensemble trained on the entire training dataset. Outliers outside of the boxplots' whiskers are not shown. The colors indicate groups of wind directions. Pastel colors in the legend indicated by an "a" stand for anticyclonic, dark colors indicated with a "c" for cyclonic circulation. Bold class name on the x-axis indicate a significant signal-to-noise ratios based on multiple testing with a significance level of 5 %.

References

Aalbers, E. E., Lenderink, G., van Meijgaard, E. and van den Hurk, B. J. J. M. (2018). Local-scale changes in mean and heavy precipitation in western europe, climate change or internal variability?, *Climate Dynamics* **50**(11-12): 4745–4766.

Beck, C., Jacobeit, J. and Jones, P. D. (2007). Frequency and within-type variations of large-scale circulation types and their effects on low-frequency climate variability in central europe since 1780, *International Journal of Climatology* **27**(4): 473–491.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal statistical society: series B (Methodological)* **57**(1): 289–300.

Cannon, A. J. (2020). Reductions in daily continental-scale atmospheric circulation biases between generations of global climate models: Cmip5 to cmip6, *Environmental Research Letters* **15**(6): 064006.

Cawley, G. C. and Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation, *Journal of Machine Learning Research* **11**: 2079–2107.

Deser, C., Phillips, A., Bourdette, V. and Teng, H. (2012). Uncertainty in climate change projections: the role of internal variability, *Climate Dynamics* **38**(3-4): 527–546.

Drücke, J., Borsche, M., James, P., Kaspar, F., Pfeifroth, U., Ahrens, B. and Trentmann, J. (2021). Climatological analysis of solar and wind energy in germany using the grosswetterlagen classification, *Renewable Energy* **164**: 1254–1266.

Hawkins, E. and Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change, *Climate Dynamics* **37**(1-2): 407–418.

Herrera-Lormendez, P., Mastrantonas, N., Douville, H., Hoy, A. and Matschullat, J. (2021). Synoptic circulation changes over central europe from 1900 to 2100: Reanalyses and coupled model intercomparison project phase 6, *International Journal of Climatology* .

Hess, P. and Brezowsky, H. (1952). Katalog der großwetterlagen europas. Ber Dt Wetterd in der USZone 33.

Hoffmann, P. and Spekat, A. (2021). Identification of possible dynamical drivers for long-term changes in temperature and rainfall patterns over europe, *Theoretical and Applied Climatology* **143**(1-2): 177–191.

Hoy, A., Hänsel, S. and Maugeri, M. (2020). An endless summer: 2018 heat episodes in europe in the context of secular temperature variability and change, *International Journal of Climatology* **40**(15): 6315–6336.

Hoy, A., Jaagus, J., Sepp, M. and Matschullat, J. (2013). Spatial response of two european atmospheric circulation classifications (data 1901–2010), *Theoretical and Applied Climatology* **112**(1-2): 73–88.

Huguenin, M. F., Fischer, E. M., Kotlarski, S., Scherrer, S. C., Schwierz, C. and Knutti, R. (2020). Lack of change in the projected frequency and persistence of atmospheric circulation types over central europe, *Geophysical Research Letters* **47**(9).

Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T. and Yang, H. (2019). Machine learning and artificial intelligence to aid climate change research and preparedness, *Environmental Research Letters* **14**(12): 124007.

Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselý, J. and Tveito, O. E. (2008). Classifications of atmospheric circulation patterns: recent advances and applications, *Annals of the New York Academy of Sciences* **1146**: 105–152.

James, P. M. (2006a). An assessment of european synoptic variability in hadley centre global environmental models based on an objective classification of weather regimes, *Climate Dynamics* **27**(2-3): 215–231.

James, P. M. (2006b). An objective classification method for hess and brezowsky grosswetterlagen over europe, *Theoretical and Applied Climatology* **88**(1): 17–42.

Jenkinson, A. and Collison (1977). *An initial climatology of gales over the North Sea*, Bracknell: Meteorological Office. Synoptic Climatology Branch Memorandum No. 62.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* .

Kučerová, M., Beck, C., Philipp, A. and Huth, R. (2017). Trends in frequency and persistence of atmospheric circulation types over europe derived from a multitude of classifications, *International Journal of Climatology* **37**(5): 2502–2521.

Kurth, T., Zhang, J., Satish, N., Mitliagkas, I., Racah, E., Patwary, M. A., Malas, T., Sundaram, N., Bhimji, W., Smorkalov, M., Deslippe, J., Shiryaev, M., Sridharan, S., Prabhat and Dubey, P. (2017). Deep learning at 15pf: Supervised and semi-supervised classification for scientific data.
URL: *http://arxiv.org/pdf/1708.05256v1*

Lakshminarayanan, B., Pritzel, A. and Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles, *Advances in neural information processing systems* **30**.

Lewis, D. D., Schapire, R. E., Callan, J. P. and Papka, R. (1996). Training algorithms for linear text classifiers, *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 298–306.

Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M. and Collins, W. (2016). Application of deep convolutional neural networks for detecting extreme weather in climate datasets.
URL: *http://arxiv.org/pdf/1605.01156v1*

Maher, N., Milinski, S. and Ludwig, R. (2021). Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble, *Earth System Dynamics* **12**(2): 401–418.

Minářová, J., Müller, M., Clappier, A., Hänsel, S., Hoy, A., Matschullat, J. and Kašpar, M. (2017). Duration, rarity, affected area, and weather types associated with extreme precipitation in the ore mountains (erzgebirge) region, central europe, *International Journal of Climatology* **37**(12): 4463–4477.

Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y. and Ludwig, R. (2019). Detecting climate change effects on vb cyclones in a 50–member single–model ensemble using machine learning, *Geophysical Research Letters* **46**(24): 14653–14661.

Petrow, T., Zimmer, J. and Merz, B. (2009). Changes in the flood hazard in germany through changing frequency and persistence of circulation patterns, *Natural Hazards and Earth System Sciences* **9**(4): 1409–1423.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L. and Fisher, M. (2016). Era-20c: An atmospheric reanalysis of the twentieth century, *Journal of Climate* **29**(11): 4083–4097.

Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., Prabhat and Pal, C. (2016). Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events.
URL: *http://arxiv.org/pdf/1612.02095v2*

Ringer, M. A., Martin, G. M., Greeves, C. Z., Hinton, T. J., James, P. M., v. d. Pope, Scaife, A. A., Stratton, R. A., Inness, P. M., Slingo, J. M. and Yang, G.-Y. (2006). The physical properties of the atmosphere in the new hadley centre global environmental model (hadgem1). part ii: Aspects of variability and regional climate, *Journal of Climate* **19**(7): 1302–1326.

Snoek, J., Larochelle, H. and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms, *Advances in neural information processing systems* **25**.

Sulikowska, A. and Wypych, A. (2020). How unusual were june 2019 temperatures in the context of european climatology?, *Atmosphere* **11**(7): 697.

Sýkorová, P. and Huth, R. (2020). The applicability of the hess–brezowsky synoptic classification to the description of climate elements in europe, *Theoretical and Applied Climatology* **142**(3-4): 1295–1309.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P. and Woollen, J. (2005). The era-40 re-analysis, *Quarterly Journal of the Royal Meteorological Society* **131**(612): 2961–3012.
URL: *https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.04.176*

Vautard, R., Yiou, P., Otto, F., Stott, P., Christidis, N., van Oldenborgh, G. J. and Schaller, N. (2016). Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events, *Environmental Research Letters* **11**(11): 114009.

Verdecchia, M., Visconti, G., D'Andrea, F. and Tibaldi, S. (1996). A neural network approach for blocking recognition, *Geophysical Research Letters* **23**(16): 2081–2084.

Werner, P. C. and Gerstengarbe, F. W. (2010). *Katalog der Großwetterlagen Europas (1881-2009) nach Paul Hess und Helmut Brezowsky*, PIK Report. Potsdam. Nr. 119, 7. Aufl.

Woollings, T., Hannachi, A. and Hoskins, B. (2010). Variability of the North Atlantic eddy-driven jet stream, *Journal of the Royal Meteorological Society* **136**: 856–868.

Wyser, K., Koenigk, T., Fladrich, U., Fuentes-Franco, R., Karami, M. P. and Kruschke, T. (2021). The smhi large ensemble (smhi-lens) with ec-earth3. 3.1, *Geoscientific Model Development* **14**(7): 4781–4796.

3 Conclusion

This thesis investigates dynamic drivers of different regional hydro-meteorological extreme events by classifying them in large climate model ensembles using machine learning techniques. The conclusion of this dissertation consists of two parts: firstly, the answers to the research questions raised in chapter 1.4, and secondly, recommendations for machine learning applications in climate pattern recognition tasks, which originate from experiences throughout the work on this thesis.

3.1 Answers to the Research Questions

In the following, the key findings of the four presented publications are summarized by answering the research questions posed in chapter 1.4. The research questions are related to two overarching research questions, which cover all four publications and span over the entire thesis. The overarching research questions are addressed at the end. The subordinated research questions are divided into three sections, with each section relating to one atmospheric feature or set of atmospheric features: Vb-cyclones, pressure-driven channeling, and the 29 circulation types called *Großwetterlagen*. The case of circulation types is covered by two publications, whereby publication III is a preliminary study for paper IV and focuses on six selected circulation types that drive European heat and drought extremes.

Vb-Cyclones

Q1.1: Can neural networks for pattern recognition detect cut-off low related Vb-cyclones in regional climate model data sets? Can this machine learning approach be combined with the established meteorological procedure of cyclone tracking?
Paper I shows that the machine learning technique of neural networks can be a useful tool for the detection of cut-off low related Vb-cyclones in climate models, especially for analyzing large ensembles. In combination with a tracking procedure, neural

networks can help to efficiently filter large data sets for the atmospheric situations of interest. The detailed tracking procedure post-processes the network output and ensures that the detected Vb-cyclones fulfill the meteorological criteria for their definition. This two-step procedure with a combination of machine learning and a classical meteorological approach ensures a high accuracy in Vb-cyclone detection. The classification results of paper I state that 94.6 % of them are correctly identified.

Q2.1: How does climate change under the RCP8.5 scenario affect the frequency, seasonality and precipitation intensity of cut-off low related Vb-cyclones? What role does internal climate variability play?

The findings of paper I reveal that in the CRCM5-LE, which is driven by the boundary conditions of the *Canadian Earth System Model version 2 Large Ensemble (CanESM2-LE)* and the RCP8.5 scenario as external forcing, the absolute frequency of Vb-cyclones per year does not change significantly. The seasonality of Vb occurrence, however, is subject to strong changes. Summer Vb-cyclones decrease by -51.8 % (median) when comparing the far future (2070-2099) to the reference period (1980-2009). This supports the findings by Nissen et al. (2013), who find a decrease in Vb-cyclone occurrence throughout the summer half-year even though using a different model and a less strong forcing scenario. Spring Vb-cyclones, on the contrary, occur considerably more often with a median of +73.4 %. This leads to a shift of the peak of maximum Vb-cyclone occurrence throughout the year from summer, the season associated with the highest precipitation intensities, to spring. The spread of internal variability in Vb-cyclone occurrence is large (at its highest in spring in future it is one event per year, while having an average probability of occurrence of also one event per year). Still, the seasonal frequency changes exceed the noise of internal variability and have signal-to-noise ratios larger than 1. With progressing climate change in the CRCM5-LE, the precipitation intensities of Vb-cyclones increase in all seasons, but remain highest in the warm season of summer.

Pressure Driven Channeling

Q1.2: Can deep learning help to identify the large-scale atmospheric driver of mixed precipitation in Montréal in regional and global climate models? How can this deep learning approach be combined with meteorological domain knowledge?

The findings of paper II illustrate that deep learning can achieve decent accuracy in classifying the large-scale atmospheric driver of mixed precipitation in the Montréal area. The combination of the CNN with a temperature and precipitation condition in a two-staged approach allows to include meteorological domain knowledge about a thermodynamic precondition for mixed precipitation occurrence. This notably increases the classification accuracy of the approach by improving the Matthews correlation coefficient by + 21 %. In the end, the trained CNN reaches a recall of 83.8 % and a precision of 72.4 % on the test set. The Matthews correlation coefficient of 0.77 indicates an overall decent performance.

Circulation Types

Q1.3: Can deep learning help to study important atmospheric drivers of heat and drought in large climate ensembles?

The workshop paper of publication III introduces a deep learning classifier that identifies six selected circulation types that trigger European heat and drought events. On these selected circulation types, the classifier reaches an overall accuracy of 60 % and a macro F1-score of 0.38. This deep learning classifier based on a CNN architecture can help to detect important atmospheric circulation patterns that are associated with these extreme events in large ensembles of climate models. Thus it can be one helpful step towards a better understanding of the dynamic drivers of heat and drought in Europe.

Q1.4: Can deep learning provide a more accurate classification method of Hess & Brezowsky's circulation types over Europe than the state-of-the-art? What uncertainties are involved in the deep learning approach?

The introduced deep learning classifier of paper IV is capable of classifying all 29 cir-

culation types from the subjective catalog by Hess & Brezowsky. It reaches an overall accuracy of 41.1 % on the test set during nested cross-validation. The comparison with the state-of-the-art method by James (2007) on ERA-40 reanalysis reveals that the deep learning based method reaches higher classification accuracies for 22 of 29 classes and thus has a higher performance in reproducing the original catalog. Due to its complexity, the deep learning procedure itself is subject to uncertainties. Paper IV quantifies the network's uncertainty due to weight initializations by generating a large *deep ensemble* of 30 networks with different random weight initializations in the beginning of the learning process. The network's uncertainty is addressed by using the weighted average of the deep ensemble to make class predictions during inference on new data.

Q2.2: How does climate change affect the frequency distribution of Hess & Brezowsky's circulation types over Europe in the Swedish Meteorological and Hydrological Institute - Large Ensemble (SMHI-LENS), a SMILE from the CMIP6 generation? What is the role of internal climate variability?

Employing the SMHI-LENS ensemble under the SSP37.0 scenario, the findings of paper IV reveal overall small absolute frequency changes of ± 5 days for most circulation types when comparing the far future (2071-2100) with a reference period (1991-2020). This is in line with the findings by Huguenin et al. (2020). For circulation types that occur only rarely, this can still mean high relative frequency changes of around ± 50 %. Generally speaking, south-easterly circulations are decreasing in both summer and winter; northerlies are decreasing in summer, while north-easterly circulations show a summer increasing trend. The spread of internal variability is considerable (with up to 18 days per year for the most frequent circulation type). The study uses the SMILE of the SMHI-LENS and the signal-to-noise ratio to discriminate between climate change and the noise of internal variability. Using a 5 %-significance level, the results show significant frequency changes for 20 out of the 29 circulation types (69 %), despite the high spread of internal climate variability.

Overarching Research Questions

Q1: Can neural networks facilitate the investigation of atmospheric drivers of regional hydro-meteorological extreme events by identifying them in large climate model ensembles?

This thesis illustrates with four publications on three atmospheric features or set of features that the machine learning technique of neural networks can reach adequate accuracies in classifying atmospheric patterns and can thus help to detect the atmospheric drivers of various different regional hydro-meteorological extreme events in climate model data. With this, neural networks can support the investigation of the dynamic component of climate change impacts on regional extremes. The benefit of neural networks comes to bear especially when working with large data sets and using neural networks as efficient filtering tool. This is especially important in the era of large ensembles and ever-growing volume of climate model data sets.

Q2: How does climate change influence the occurrence of atmospheric drivers of regional hydro-meteorological extreme events? What role does internal climate variability play?

As the results of the climate change studies in paper I and IV of this thesis show, internal climate variability is an important source of uncertainty for the atmospheric drivers of regional hydro-meteorological extreme events. The spread of internal climate variability is illustrated in form of boxplots in Figure 2 and 3 of paper I and Figure 4 of paper IV and is considerable for Vb-cyclones and circulation types. Using SMILES, climate trends can be separated from the noise of internal climate variability. This reveals effects of climate change on the frequency with which the atmospheric drivers occur. In the case of Vb-cyclones this primarily includes a shift in seasonality, while for circulation types 20 of 29 classes (69 %) show significant absolute frequency changes for the entire year.

3.2 Recommendations for machine learning applications in climate pattern recognition

There are three main recommendations for best practices in machine learning applications in the area of climate pattern recognition that are set up based on the experiences and the orientation of this thesis.

Integrating meteorological domain knowledge

The first recommendation is targeted at the integration of meteorological domain knowledge into the design of machine learning approaches. This idea was pursued throughout this thesis. For paper I and II this was implemented in form of a two-staged approach, in which neural networks were combined with a meteorological processing step, e.g., cyclone tracking (paper I) or post-processing with a temperature and precipitation criterion (paper II). For paper III and IV a subjective catalog of circulation types was taken as ground-truth, which was generated by experts based on their synoptic domain knowledge. From a computer science perspective, the ideal state often aimed for is the integration of domain knowledge into an end-to-end learning (Glasmachers 2017), where the entire learning process is captured by one machine learning model without the need for post-processing. An example in the case of circulation types is given: the transition smoothing step, which insures that a circulation type fulfills its meteorological definition in that it lasts for at least three days, could be realized directly during classification by a deep hidden Markov model (Mittermeier et al. 2021b). This would be an end-to-end learning process without the need for post-processing. However, it is important not to carry out an end-to-end learning at the expense of domain knowledge. In case of the tracking procedure applied in paper I, for example, there are good arguments for doing an extra post-processing step: this way an well-established procedure can be applied, which is comparable with other research studies, and follows a physically-based approach. A promising approach of combining neural networks with climate domain knowledge differing from the strategy followed in this thesis, is the integration of machine learning in existing climate models, e.g.,

by replacing the parameterization of clouds in climate models by a machine learning model as mentioned in chapter 1.2.2 (Rolnick et al. 2019).

Choosing a balanced and transparent test set

The second recommendation is about choosing a balanced test set and making the test set choice transparent. A test requires to be representative of the wider population (Lones 2021). For climate pattern recognition tasks there are often challenges introduced by the time series structure of the data and strongly imbalanced class distributions in the case of extreme event detection. These special features are to be be taken into account when selecting the test set. It is highly recommended to make the test set choice transparent by presenting how the test set is derived and composed (Mittermeier et al. 2021a).

Acknowledging the elaborative data processing and training process

The third recommendation is to acknowledge that the elaborative and time consuming part in classifying atmospheric patterns using machine learning lies in the step of designing the approach, processing the data and training the network. This is especially the case when the number of training examples is particularly high (as in paper II). The argument of efficiency in terms of machine learning is related to the application of the final product - the trained network - to new data. This step is quick and of low computational cost. The step of establishing the procedure and enhancing the network performance, however, can be time consuming and potentially requires hands-on work in analyzing the weak points and testing different strategies for possible performance achievement. If the trained network is used only for one specific research scope, this raises the question of efficiency in respect to the research workflow. A solution can be to design a deep learning classifier for the identification of several different atmospheric features at once. This way, the dynamic drivers of different extreme events can be analyzed and one deep learning classifier can be used to address various research questions. This is a major benefit of paper III and IV as they focus on a set of circulation types with 29 different classes that are associated with several different extreme events.

References

Aalbers, E. E., Lenderink, G., van Meijgaard, E. & van den Hurk, B. J. J. M. (2018): Local-scale changes in mean and heavy precipitation in western europe, climate change or internal variability? *Climate Dynamics*, 50(11-12), 4745–4766. DOI: 10.1007/s00382-017-3901-9.

Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K. & Taha, K. (2015): Efficient machine learning for big data: A review. *Big Data Research*, 2(3), 87–93. DOI: 10.1016/j.bdr.2015.04.001.

Barry, R. G. (2005): Synoptic climatology. In: Oliver, J. E. (ed.), Encyclopedia of World Climatology, Springer, Bodmin, Cornwall, 700–705.

Benestad, R., Sillmann, J., Thorarinsdottir, T. L., Guttorp, P., Mesquita, M. d. S., Tye, M. R., Uotila, P., Maule, C. F., Thejll, P., Drews, M. & Parding, K. M. (2017): New vigour involving statisticians to overcome ensemble fatigue. *Nature Climate Change*, 7(10), 697–703. DOI: 10.1038/nclimate3393.

Bertola, M., Viglione, A., Lun, D., Hall, J. & Blöschl, G. (2020): Flood trends in europe: are changes in small and big floods different? *Hydrology and Earth System Sciences*, 24(4), 1805–1822.

Bracken, L. & Oughton, E. (2009): Interdisciplinarity within and beyond geography: introduction to special section. *Area*, 41(4), 371–373.

Brunner, M. I., Swain, D. L., Wood, R. R., Willkofer, F., Done, J. M., Gilleland, E. & Ludwig, R. (2021): An extremeness threshold determines the regional response of floods to changes in rainfall extremes. *Communications Earth & Environment*, 2(1). DOI: 10.1038/s43247-021-00248-x.

Cawley, G. C. & Dorling, S. R. (1996): Reproducing a subjective classification scheme for atmospheric circulation patterns over the united kingdom using a neural network. In: International Conference on Artificial Neural Networks, Springer, Berlin, Heidelberg, 281–286.

Chattopadhyay, A., Hassanzadeh, P. & Pasha, S. (2018): A test case for application of convolutional neural networks to spatio-temporal climate data: Re-identifying clustered weather patterns. `http://arxiv.org/pdf/1811.04817v1`.

Chattopadhyay, A., Nabizadeh, E. & Hassanzadeh, P. (2020): Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of advances in modeling earth systems*, 12(2), e2019MS001958. DOI: 10.1029/2019MS001958.

Chen, Y., Moufouma-Okia, W., Masson-Delmotte, V., Zhai, P. & Pirani, A. (2018): Recent progress and emerging topics on weather and climate extremes since the fifth assessment report of the intergovernmental panel on climate change. *Annual Review of Environment and Resources*, 43(1), 35–59. DOI: 10.1146/annurev-environ-102017-030052.

Deser, C., Phillips, A., Bourdette, V. & Teng, H. (2012): Uncertainty in climate change projections: the role of internal variability. *Climate Dynamics*, 38(3-4), 527–546. DOI: 10.1007/s00382-010-0977-x.

Dunn, R. J., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., Lippmann, T., Allan, R., Aguilar, E., Barry, A. A. et al. (2020): Development of an updated global land in situ-based data set of temperature and precipitation extremes: Hadex3. *Journal of Geophysical Research: Atmospheres*, 125(16), e2019JD032263.

Farinosi, F., Dosio, A., Calliari, E., Seliger, R., Alfieri, L. & Naumann, G. (2020): Will the paris agreement protect us from hydro-meteorological extremes? *Environmental Research Letters*, 15(10), 104037. DOI: 10.1088/1748-9326/aba869.

Gardner, M. W. & Dorling, S. R. (1998): Artificial neural networks (the multilayer perceptron) — a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15), 2627–2636.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G. & Yacalis, G. (2018): Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11), 5742–5751. DOI: 10.1029/2018GL078202.

Giorgi, F. & Gutowski, W. J. (2015): Regional dynamical downscaling and the cordex initiative. *Annual Review of Environment and Resources*, 40(1), 467–490. DOI: 10.1146/annurev-environ-102014-021217.

Glasmachers, T. (2017): Limits of end-to-end learning. In: Asian Conference on Machine Learning, PMLR, 17–32.

Goodfellow, I., Bengio, Y. & Courville, A. (2016): Deep Learning. MIT Press. `http://www.deeplearningbook.org`.

Grotjahn, R., Black, R., Leung, R., Wehner, M. F., Barlow, M., Bosilovich, M., Gershunov, A., Gutowski, W. J., Gyakum, J. R., Katz, R. W., Lee, Y.-Y., Lim, Y.-K. & Prabhat (2016): North american extreme temperature events and related large scale meteorological patterns: a review of statistical methods, dynamics, modeling, and trends. *Climate Dynamics*, 46(3-4), 1151–1184. DOI: 10.1007/s00382-015-2638-6.

Hagen, M. T., Demuth, H. B., Beale, M. H. & De Jesus, O. (2014): Neural network design. E-book: `http://hagan.okstate.edu/nnd.html`.

Hawkins, E. & Sutton, R. (2011): The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, 37(1-2), 407–418. DOI: 10.1007/s00382-010-0810-6.

Horton, D. E., Johnson, N. C., Singh, D., Swain, D. L., Rajaratnam, B. & Diffenbaugh, N. S. (2015): Contribution of changes in atmospheric circu-

lation patterns to extreme temperature trends. *Nature*, 522(7557), 465–469. DOI: 10.1038/nature14550.

Huguenin, M. F., Fischer, E. M., Kotlarski, S., Scherrer, S. C., Schwierz, C. & Knutti, R. (2020): Lack of change in the projected frequency and persistence of atmospheric circulation types over central europe. *Geophysical Research Letters*, 47(9). DOI: 10.1029/2019GL086132.

Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T. & Yang, H. (2019): Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters*, 14(12), 124007. DOI: 10.1088/1748-9326/ab4e55.

IPCC (2013): Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

James, P. M. (2007): An objective classification method for hess and brezowsky grosswetterlagen over europe. *Theoretical and Applied Climatology*, 88(1-2), 17–42. DOI: 10.1007/s00704-006-0239-3.

Knüsel, B., Zumwald, M., Baumberger, C., Hirsch Hadorn, G., Fischer, E. M., Bresch, D. N. & Knutti, R. (2019): Applying big data beyond small problems in climate research. *Nature Climate Change*, 9(3), 196–202. DOI: 10.1038/s41558-019-0404-1.

Kurth, T., Zhang, J., Satish, N., Mitliagkas, I., Racah, E., Patwary, M. A., Malas, T., Sundaram, N., Bhimji, W., Smorkalov, M., Deslippe, J., Shiryaev, M., Sridharan, S., Prabhat & Dubey, P. (2017): Deep learning at 15pf: Supervised and semi-supervised classification for scientific data. `http://arxiv.org/pdf/1708.05256v1`.

Lagerquist, R., McGovern, A. & & Gagne II, D. J. (2019): Deep learning for spatially explicit prediction of synoptic-scale fronts. *Weather and Forecasting*, 34(4), 1137–1160. DOI: 10.1175/WAF-D-18-0183.s1.

LeCun, Y. (1989): Generalization and network design strategies. *Connectionism in perspective*, 19(143-155), 18.

LeCun, Y., Bengio, Y. & Hinton, G. (2015): Deep learning. *Nature*, 521(7553), 436–44.

Leduc, M., Mailhot, A., Frigon, A., Martel, J.-L., Ludwig, R., Brietzke, G. B., Giguère, M. & Brissette, F., Turcotte, R., Braun, M., Scinocca, J. (2019): The climex project: A 50-member ensemble of climate change projections at 12-km resolution over europe and northeastern. *American Meteorological Society*, 663–693. DOI: 10.1175/JAMC-D-18-0021.s1.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R. & Hawkins, E. (2020): Partitioning climate projection uncertainty with multiple large ensembles and cmip5/6. *Earth System Dynamics*, 11(2), 491–508. DOI: 10.5194/esd-11-491-2020.

Liu, W., Cheng, H. & Han, X. (2022): Rebuilding geography for the 21st century through disciplinary reunification and social engagement. *Environment and Planning F*, 26349825221082162.

Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M. & Collins, W. (2016): Application of deep convolutional neural networks for detecting extreme weather in climate datasets. `http://arxiv.org/pdf/1605.01156v1`.

Lones, M. A. (2021): How to avoid machine learning pitfalls: a guide for academic researchers. School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, Scotland, UK. `http://arxiv.org/pdf/2108.02497v1`.

Maher, N., Milinski, S. & Ludwig, R. (2021): Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth System Dynamics*, 12(2), 401–418.

Matthews, B. W. (1995): Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2), 442–451. DOI: 10.1016/0005-2795(75)90109-9.

McCulloch, W. S. & Pitts, W. (1943): A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115–133.

McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T. & Williams, J. K. (2017): Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bulletin of the American Meteorological Society*, 98(10), 2073–2090. DOI: 10.1175/BAMS-D-16-0123.1.

Mittermeier, M., Bresson, É., Paquin, D. & Ludwig, R. (2021a): A deep learning approach for the identification of long-duration mixed precipitation in montréal (canada). *Atmosphere-Ocean*, 1–12. DOI: 10.1080/07055900.2021.1992341.

Mittermeier, M., Weigert, M. & Rügamer, D. (2021b): Identifying the atmospheric drivers of drought and heat using a smoothed deep learning approach. NeurIPS workshop 2021: Tackling Climate Change with Machine Learning. *arXiv preprint*. `http://arXiv.org/pdf/2111.05303`.

Molina, M. J., Gagne, D. J. & Prein, A. F. (2021): A benchmark to test generalization capabilities of deep learning methods to classify severe convective storms in a changing climate. *Earth and Space Science*, 8(9). DOI: 10.1029/2020EA001490.

Muszynski, G., Kashinath, K., Kurlin, V. & Wehner, M. (2019): Topological data analysis and machine learning for recognizing atmospheric river patterns in large climate datasets. *Geoscientific Model Development*, 12(2), 613–628. DOI: 10.5194/gmd-12-613-2019.

Nissen, K. M., Ulbrich, U. & Leckebusch, G. C. (2013): Vb cyclones and associated rainfall extremes over central europe under present day and climate change conditions. *Meteorologische Zeitschrift*, 22(6), 649–660. DOI: 10.1127/0941-2948/2013/0514.

O'Shea, K. & Nash, R. (2015): An introduction to convolutional neural networks. `http://arxiv.org/pdf/1511.08458v2`.

Oueslati, B., Yiou, P. & Jézéquel, A. (2019): Revisiting the dynamic and thermodynamic processes driving the record-breaking january 2014 precipitation in the southern uk. *Scientific reports*, 9(1), 2859. DOI: 10.1038/s41598-019-39306-y.

Overpeck, J. T., Meehl, G. A., Bony, S. & Easterling, D. R. (2011): Climate data challenges in the 21st century. *Science (New York, N.Y.)*, 331(6018), 700–702. DOI: 10.1126/science.1197869.

Pfahl, S., O'Gorman, P. A. & Fischer, E. M. (2017): Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Climate Change*, 7(6), 423–427.

Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., Prabhat, M. & Pal, C. (2017): Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. Advances in Neural Information Processing Systems 30 (NIPS 2017). `https://proceedings.neurips.cc/paper/2017/file/519c84155964659375821f7ca576f095-Paper.pdf`.

Racah, E., Beckham, C., Maharaj, T., Prabhat, M. & Pal, C. (2016): Semi-supervised detection of extreme weather events in large climate datasets. ICLR conference. `http://arxiv.org/pdf/1612.02095v1`.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-

Brown, A., Luccioni, A., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C., Ng, A. Y., Hassabis, D., Platt, J. C., Creutzig, F., Chayes, J. & Bengio, Y. (2019): Tackling climate change with machine learning. http://arxiv.org/pdf/1906.05433v2.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986): Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.

Sandu, I., Bechtold, P., Beljaars, A., Bozzo, A., Pithan, F., Shepherd, T. G. & Zadra, A. (2016): Impacts of parameterized orographic drag on the northern hemisphere winter circulation. *Journal of advances in modeling earth systems*, 8(1), 196–211. DOI: 10.1002/2015MS000564.

Schnase, J. L., Lee, T. J., Mattmann, C. A., Lynnes, C. S., Cinquini, L., Ramirez, P. M., Hart, A. F., Williams, D. N., Waliser, D., Rinsland, P., Webster, W. P., Duffy, D. Q., McInerney, M. A., Tamkin, G. S., Potter, G. L. & Carriere, L. (2016): Big data challenges in climate science: Improving the next-generation cyberinfrastructure. *IEEE Geoscience and Remote Sensing Magazine*, 4(3), 10–22. DOI: 10.1109/MGRS.2015.2514192.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M. & Zhou, B. (2021): Weather and climate extreme events in a changing climate. In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R. & Zhou, B. (eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change., Cambridge University Press. In Press.

Shepherd, T. G. (2014): Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, 7(10), 703–708. DOI: 10.1038/ngeo2253.

Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L. V., Hegerl, G., Seneviratne, S. I., Vautard, R., Zhang, X. & Zwiers, F. W. (2017): Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather and Climate Extremes*, 18, 65–74. DOI: 10.1016/j.wace.2017.10.003.

Silva, T. C. & Zhao, L. (2016): Machine learning in complex networks, volume 1. Springer.

Suarez-Gutierrez, L., Müller, W. A., Li, C. & Marotzke, J. (2020): Dynamical and thermodynamical drivers of variability in european summer heat extremes. *Climate Dynamics*, 54(9-10), 4351–4366. DOI: 10.1007/s00382-020-05233-2.

Van Bebber, W. (1891): Die Zugstrassen der barometrischen Minima nach den Bahnenkarten der deutschen Seewarte für den Zeitraum 1875-1890. *Archiv der deutschen Seewarte*, 8, 361–366.

Verdecchia, M., Visconti, G., D'Andrea, F. & Tibaldi, S. (1996): A neural network approach for blocking recognition. *Geophysical Research Letters*, 23(16), 2081–2084. DOI: 10.1029/96GL01810.

Wood, R. R., Lehner, F., Pendergrass, A. G. & Schlunegger, S. (2021): Changes in precipitation variability across time scales in multiple global climate model large ensembles. *Environmental Research Letters*, 16(8), 084022. DOI: 10.1088/1748-9326/ac10dd.

Wood, R. R. & Ludwig, R. (2020): Analyzing internal variability and forced response of subdaily and daily extreme precipitation over europe. *Geophysical Research Letters*, 47(17). DOI: 10.1029/2020GL089300.