Development of a Machine Learning Model for Hail Prediction in Germany

Master's Thesis in
Meteorology and Climate Physics
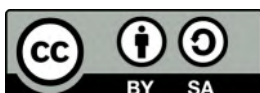by

Siyu Li

July 2023

INSTITUTE OF METEOROLOGY AND CLIMATE RESEARCH

KARLSRUHE INSTITUTE OF TECHNOLOGY (KIT)

Supervisor: Prof. Dr. Micheal Kunz
Co-supervisor: Prof. Dr. Peter Knippertz

Abstract

Hail prediction still is a major challenge in atmospheric sciences, mainly because of the complex interaction of various processes across a wide range of scales, limited knowledge in hail formation processes, and scarcely available direct observations. Machine learning, as a state-of-the-art approach, is promisingly seen to potentially overcome obstracles in hail prediction, thereby advancing our understanding of environmental preconditions of hailstorms. This thesis investigates the potential of using machine learning models for predicting the area affected by potential hailstorms in Germany on a daily basis. The overarching goal is to improve the accuracy and reliability of hailstorm prediction. ERA5 reanalysis data for convective parameters and radar-derived data for hail-affected area are implemented to explore the relationship between the preconvective hailstorm environment and hailstorm occurrence.

Initial analysis reveals that hail frequencies increase from the northern to the southern regions due to more favourable convective environmental conditions. This investigation also identifies potential convective parameters for hailstorm prediction such as Convective Available Potential Energy (CAPE), Bulk Wind Shear (BULKSH), CAPE $\times$ BULKSH (CAPESHEAR). Among 18 pre-selected convective parameters, seven indicators are identified as the best predictors. In particular, CAPESHEAR emerges as the best single predictive indicator.

The convolutional neural network (CNN) models used in this thesis are trained using 15 years of ERA5 reanalysis and radar-derived potential hailstorm streaks with promising results. Notably, these models outperform the other 2 reference models, climatology model and persistence model respectively, demonstrating its lower Mean Absolute Errors (MAEs) to those traditional non-machine learning models. And CNN models manage to realize the catagorical forecast of low vs. high hail-affected area, with considerably moderate Heidke skill score of 0.66.

The results highlight the added value of the machine learning model compared to conventional methods such as statistical regression. It is designed in this thesis to predict aggregated hailstorm clusters across Germany rather than isolated storms, based on the additional class activation analysis. Depending on this, the predictive performance of machine learning still remains relatively robust and needs further optimization.

The machine learning model showcases remarkable potential in predicting the area affected by radar-identified hailstorms. Even though uncertainties still persist, it is shown that machine learning combined with other predictive model formats, such as ensemble forecasting, may significantly enhance the accuracy of hailstorm prediction in the future research.

Contents

1 Introduction

Numerous studies over recent decades have shown that the probability of severe convective storms (SCSs) on the regional scale will increase with climate change in the future (Trapp et al., 2007a; Cao, 2008; Kunz et al., 2009; Seneviratne et al., 2012; Diffenbaugh et al., 2013). These SCSs are associated with various local extreme weather phenomena such as hail, wind gusts, tornadoes, and heavy precipitation. Hail as one of the most destructive natural hazards in Central Europe and other regions poses a substantial threat to the society by severely damaging buildings, infrastructures and agriculture, leading to a large amount of economic damage. The insured losses caused mainly by hail during two recent events in Germany (Kunz et al., 2018; Wilhelm et al., 2021) by a series of SCSs sum up to €2.7 billion and €0.75 billion, respectively (MunichRE, 2020). In view of the severe impact resulting from hail-related SCSs, an in-depth understanding of their favourable environmental conditions is needed, aiming to improve the accuracy of hail prediction and to help society to take preventive actions in a timely manner.

Organized convective storms such as multicells, supercells, or mesoscale convective storms (MCSs) are frequently accompanied by hail (Auer, 1972; Markowski and Richardson, 2011). Different types of observation data are essential for the most direct and accurate investigation on the characteristics of hail-produced thunderstorms. Meanwhile, direct measurements of hail has always been facing difficulties mainly because of the limited regions usually affected by hail in combination with a lack of instruments to detect hail and their sizes. Although geostationary satellite measurements can indirectly observe hail-producing clouds, they are still not able to precisely monitor hailstorm occurrence under the cloud when capturing, for example, the overshooting top of thunderstorms (Punge and Kunz, 2016). In Germany, weather radar networks with high spatial and temporal resolutions are able to monitor both the intensity and the vertical structure or other signatures of thunderstorm. However, the data suffer from several errors and inaccuracies, such as by beam shielding and attenuation effects of the signal (Puskeiler et al., 2016; Wapler, 2017; Fluck et al., 2021).

In addition, numerical model simulations, proximity soundings, or reanalysis data are widely used to estimate the occurrence or frequency of SCSs, respectively (Lee, 2002; Brooks, 2013; Púčik et al., 2015; Kunz et al., 2020). Trapp et al. (2007b), for example, offered a proof-of-concept for dynamically downscaling global climate simulations to a horizontal grid spacing of 4 km or less, at which parameterizations of deep convection are not necessary any more. Based on that, Trapp et al. (2011) estimated the climatological occurrence distribution of SCSs through a model proxy that combines the core of convective updraft with the presence of hail, damaging surface winds and tornadoes. The authors found proxy occurrences coupled with information on

the large-scale atmospheric conditions could guide the trend in observed occurrences of SCSs. However, knowledge in cloud microphysics and physical mechanisms on the convection initiation, intensification and the life cycle of SCSs has not been completely understood yet. This results from uncertainties regarding the initial conditions, the inability of numerical models to represent the underlying processes and scale-interactions reliably, and the high computational costs required to run ensembles of sophisticated, highly-resolved models over long time periods (Gensini and Mote, 2015).

Several studies have investigated the relationships between hail-related SCS and the large-scale atmospheric environment (Trapp et al., 2007a; Aran et al., 2011; Piper and Kunz, 2017; Kunz et al., 2020). A major obstacle in reliable analyses is still the difficulty of accurately predicting hail occurrence. According to Prein and Holland (2018), machine learning algorithms recently have demonstrated promising results in hail prediction with higher skill compared to the process-driven numerical forecast models (Gagne et al., 2015; Czernecki et al., 2019; Gagne II et al., 2019). Leveraging the robust data processing and information extraction capabilities of machine learning, based on its outcome-driven model technique, provides a shift in how to approach the complexities of SCS prediction, which is different to classical numerical weather prediction models. This approach can bypass the complicated physical mechanisms and evolution of hailstorms. It is also expected to look at the relationship between convective storms and their atmospheric environment from a fresh and innovative perspective, and may promise to fundamentally reshape our understanding and prediction of hail events when combined with other forecast methods.

To scrutinize relationships between environmental properties and observed hail damage, several earlier studies utilized large-scale environmental conditions from reanalysis, convective parameters derived from radiosonde data (Brooks, 2009; Mohr and Kunz, 2013), remote sensing data (Fluck et al., 2021), or a combination thereof (Punge and Kunz, 2016). The objectives of this studies were to examine the preconvective environment and to predict hail occurrence. These studies also aimed to improve the skill of hail predictions. In this study, we provide a regional assessment of hail hazard by quantifying the potentially impacted area by hailstorms using radar-identified hail tracks. Further, the study facilitates the prediction of hail occurrences leveraging environmental conditions, taking the advantage of the power of convolutional neural networks within a machine learning framework. The thesis focus on the following research questions:

1. Is there a characteristic geographic distribution of the prevailing convective environment for hailstorms in Germany?

2. Is machine learning useful for hail prediction and for better understanding relationships between ambient conditions and hailstorms? What is the predictive skill of the machine learning models?

3. Which predictors yield the best model performance, and how do they modify the accuracy of the model prediction?

4. When and why are machine learning models unable to predict hail occurrence?

The structure of this Master's thesis unfolds as follows: Chapter 2 presents theoretical concepts considered in this thesis, covering aspects of convection mechanisms, types of convective systems, radar measurements, and machine learning principles. In Chapter 3, we introduce the data used in this thesis and elucidate the methods used in considerable detail. Subsequently, the findings of the study are systematically displayed across Chapters 4–6. Chapter 4 kicks off this series with a statistical analysis of the hailstorm environment. Building upon this groundwork, Chapter 5 critically analyses the performance evaluation of the prediction model. The insights gained from this evaluation then inform the discussions in Chapter 6, where we explore the meteorological interpretations from the machine learning model, with a particular focus on identifying errors within the model predictions. The final chapter, Chapter 7, summarises the findings of this thesis and suggests future directions for this research.

2 Theoretical concepts

Hail-producing thunderstorms have a large vertical extent through the troposphere. They are formed by thermodynamic processes of lifted air driving the vertical dynamics of the atmosphere. This chapter introduces the theoretical concepts of thunderstorm and hail formation and the methods on which this thesis builds upon. The fundamentals of convection are included in Section 2.1; Section 2.2 gives a brief overview of different types of convective storms; Section 2.3 introduces the basis of radar measurement and a short introduction of machine learning algorithm is written in Section 2.4. The previous research on severe thunderstorms summarized in the end of this chapter (Section 2.5).

2.1 Convection

Convection in the atmosphere is the transfer of heat through the movement of air due to temperature differences. The vertical stratification leads to the development of the convection that influences weather patterns. Atmospheric instability, low-level moisture, and a trigger mechanism for the lifting of air parcels to the level of free convection are three necessary, but not sufficient requirements crucial for the development of deep moist convection (Markowski and Richardson, 2011). The uneven distribution of vertical temperature gradient in the atmosphere contributes to the instability of the atmosphere. Combining with the process of free convection further triggers deep convection.

2.1.1 Thermodynamic fundamentals

The three existent phases of water in the atmosphere are: solid (ice, hail, snow), liquid (raindrops, cloud droplets), vapor (water vapor). Phase changes of water vapor in the atmosphere have a great impact on the thermodynamic processes in clouds and on the atmospheric radiation in the energy budget of Earth.

The equation of state for an ideal gas (dry) is

$$p_d = \rho_d R_d T, \tag{2.1}$$

where p_d is the pressure of dry air, ρ_d is the density of the dry air, T is the temperature, and $R_d = 287.04 \, \text{Jkg}^{-1} \, \text{K}^{-1}$ is the gas constant for dry air.

The most essential law for governing the gaseous water vapor and energy transfer is the first law of thermodynamics

$$\mathrm{d}u = \delta q + \delta a \, , \tag{2.2}$$

where $\mathrm{d}u$ is the change of internal energy as a result from the variations of specific heat δq taken into or released out of a specific volume and the work variations δa from its environment. If air is compressible, δa can be expressed as:

$$\delta a = -p d\alpha \tag{2.3}$$

where α is the specific volume $(= \rho^{-1})$. δq and δa are path-independent differentials and are written with a d instead of a δ.

The variation of internal energy is determined by the temperature change, namely Joule's second law $(du = c_v dT)$. With the enthalpy $h = u + p\alpha$, Eq.(2.2) becomes

$$\mathrm{d}q = \mathrm{d}h - p\mathrm{d}\alpha \tag{2.4}$$

where $\mathrm{d}h = c_p \mathrm{d}T$, with c_p as the specifc hear capacity fro the constant pressure and an adiabatic system. The specific heating rate can also contain the heating and cooling as a result of radiation, molecular diffusion and phase change of water, can also decribes the change in internal energy, so Eq.(2.4) can re-written as:

$$\mathrm{d}q = c_p \mathrm{d}T - \alpha \mathrm{d}p. \tag{2.5}$$

If a dry air parcel is lifted and experiences a lower pressure, it expands and, consequently, cools dry-adiabatically according to the first law of thermodynamics. This quantity can be derived from Eq.(2.5), and represents cooling (warming) of an air parcel resulting from expansion (compression) as it ascends (descends) dry-adiabatically. When $\mathrm{d}q = 0$, there is no exchange of heat between the air parcel and the environment, referred to as adiabatic. Based on the assumption of hydrostatic balance, which is assumed for the derivation of the temperature gradient in the barotropic atmosphere $(dp = -\rho g dz)$, together with the equation of state for an ideal gas, yields

$$\Gamma_d \equiv -\frac{dT}{dz} = -\frac{g}{c_p}; \tag{2.6}$$

which is for mid-latitudes $\Gamma_d \approx 9.8$ K km^{-1}. This gradient is valid for dry and also moist (unsaturated) air with the specific heat capacity

$$c_{pm} = x_v c_{pv} + x_v c_{pd} \tag{2.7}$$

with indices m for the moist air, v for water vapor, d for dry air and x_v for the mass fraction of water vapor, if $c_{pv} \approx c_{pd}$ when the condensation does not occur.

2.1.2 Thermal instability of the atmosphere

Convective stability or instability of the atmosphere can be related to vertical temperature differences, which can give an idea of the potential for thunderstorm development.

1) Potential temperature:

Integrating Eq. (2.4) from p to a reference pressure p_0 and assuming $\delta q = 0$, the potential temperature equation following a *Possion's equation* is obtained:

$$\theta = T \left(\frac{p_0}{p}\right)^{\frac{R}{c_p}},\tag{2.8}$$

where θ is known as the potential temperature. This quantity is conserved for adiabatic lifting or descending of an unsaturated air parcel. It can be used to describe the stratification of an unsaturated air mass.

2) Equivalent potential temperature:

During the expanding and cooling processes of lifting unsaturated air parcels, they rise adiabatically. Compared to the dry adiabatic process, latent heat is released when air becomes saturated, reducing the cooling effect. When $\delta q = 0$ is assumed and condensation is taken into account as an internal heating process, then Eq. (2.4) can be written as:

$$dh = c_p dT + L_v dr_s - \alpha dp = 0,\tag{2.9}$$

where $L_v \approx 2.501 \times 10^6$ J kg^{-1} (temperature at 0°C) is the evaporation rate, and r_s the saturation mixing ratio. Therefore, lifting increases enthalpy with latent heat of condensation for saturated ascend in comparison to a dry case.

Then the equivalent potential temperature is obtained by integrating Eq.(2.9):

$$\theta_e \approx \theta \cdot e^{\left(\frac{r_s L_v}{c_p T}\right)}.\tag{2.10}$$

The equivalent potential temperature θ_e is the temperature an air parcel becomes when all of the water vapor has completely condensed during lifting, and when the parcel descends to the original position afterwards. The equation is simplified by the specific formula from Bolton (1980), which gives an empirical relation between θ_e, temperature and mixing ratio. In a moist-adiabatic process, θ_e is constant with height.

The motion of an air parcel in different environments:

The vertical displacement of an air parcel driven by buoyancy is determined by the thermal instability. That is, an air parcel in a statically unstable environment is accelerated from its equilibrium position by the buoyancy force in the vertical direction in cases where its temperature or density, respectively, differ from the evironment (Markowski and Richardson, 2011).

The vertical momentum equation of air parcel presumed to assess the static instability is:

$$\frac{dw}{dt} = B, \tag{2.11}$$

where w is the vertical velocity and B is the buoyancy force. Here perturbations from pressure, Coriolis force and friction are neglected. It is also assumed that the air parcel does not disturb or influence the atmosphere environment during lifting. These assumption made above (Eq. 2.11) are referred to as parcel theory (Markowski and Richardson, 2011).

The buoyancy force B can be expressed by the environmental temperature $\overline{T}$ and the temperature T of an individual air parcel:

$$B \approx g\frac{T - \overline{T}}{\overline{T}} \tag{2.12}$$

neglecting pressure perturbations, water vapor, or condensate on the density of an air parcel. A vertically displaced parcel rises from an initial level z_0, where the parcel is in the equilibrium with its environment (i.e., $\overline{T}(z_0) = T_0 = $ T). Combining Equations (2.11) and (2.12) leads to

$$\frac{d^2\Delta z}{dt^2} = g\frac{T - \overline{T}}{\overline{T}}, \tag{2.13}$$

with Δz as the distance of the vertically displaced air parcel started from the initial level, thus $z = z_0 + \Delta z$ at some time steps late. The temperature of the parcel at its new position derived from a first-order Taylor series gives

$$T = T_0 - \Gamma_{par}\Delta z, \tag{2.14}$$

where $\Gamma_{par} = -\partial T/\partial z$ is the parcel's lapse rate. If the parcel is saturated (unsaturated) during ascent, then its lapse rate becomes the dry adiabatic lapse rate, ($\Gamma_{par} = \Gamma_d$) otherwise the moist adiabatic lapse rate ($\Gamma_{par} = \Gamma_m$). The environmental temperature at the new position is (from a first-order Taylor series approximation)

$$\overline{T} = T_0 - \gamma\Delta z, \tag{2.15}$$

where $\gamma = \partial\overline{T}/\partial z$ is the environmental lapse rate. Therefore, based on Eq. (2.13) and Eq. (2.14), Eq. (2.12) can be written as

$$\frac{d^2\Delta z}{dt^2} = -g\frac{\Gamma_{par} - \gamma}{T_0 - \gamma\Delta z}\Delta z. \tag{2.16}$$

For the reasonable estimation, due to $T_0 \gg \gamma\Delta z$ when vertical movement is quite small regarding the continuity of atmosphere, then Eq. 2.15 can be modified as

$$\frac{d^2\Delta z}{dt^2} + \frac{g}{T_0}(\Gamma_{par} - \gamma)\Delta z = 0. \tag{2.17}$$

Equation (2.17) is a second-order partial differential equation, which has a standard solution:

$$\Delta z(t) = C_1 e^{i[\frac{g}{T_0}(\Gamma_{par}-\gamma)]^{1/2}t} + C_2 e^{-i[\frac{g}{T_0}(\Gamma_{par}-\gamma)]^{1/2}t} \ . \tag{2.18}$$

Here C_1 and C_2 are constants which are related to the magnitude and direction of the orginal displacement of air parcel.

For $\gamma < \Gamma_{par}$, the real part of Eq. (2.18), $i[\frac{g}{T_0}(\Gamma_{par}-\gamma)]^{1/2}$, is imaginary and can be written as

$$\Delta z(t) = C \left\{ \cos\left[\frac{g}{T_0}(\Gamma_{par}-\gamma)\right]^{1/2}t \right\}, \tag{2.19}$$

which means that the parcel oscillates around its initial position z_0, which is regarded as a *statically stable* environmental lapse rate.

For $\gamma > \Gamma_{par}$, the part $i[\frac{g}{T_0}(\Gamma_{par}-\gamma)]^{1/2}$ is real and turned to be like below, as t becomes large, then Eq. (2.18) turns out to be:

$$\Delta z(t) = C_1 e^{[\frac{g}{T_0}(\gamma-\Gamma_{par})]^{1/2}t}, \tag{2.20}$$

indicating the instability of exponentially increased displacement of the air parcel, therefore Eq. (2.20) can not know where the air parcel stops rising and is only valid for relatively small Δz, as the infinite displacement of the parcel is not assumed with the linear profile of environmental temperature.

According to Haurwitz (1941), four situations of static instability are distinguished here.

1) An environmental lapse rate, for which $\gamma > \Gamma_d$, is so-called *absolutely unstable*:

the lifted air parcel is warmer and less dense compared to its environment, therefore will be accelerated away from its initial position depending on positive buoyancy.

2) When $\gamma < \Gamma_m$ the environmental lapse rate, namely *absolutely stable*:

the lifted air parcel is colder (higher density) than the environment, thus returns to its initial position.

3) When $\Gamma_m < \gamma < \Gamma_d$, the environmental lapse rate is *conditionally unstable*:

a stable environment with respect to an unsaturated lifted air parcel, which become unstable in case of saturated lifting. The air parcel can keep rising further after becoming saturated under such conditions. This kind of instability therefore is most important for thunderstorm development.

4) When $\gamma = \Gamma_d$ or $\gamma = \Gamma_m$, the environmental lapse rate is dry-*neutral* or moist-*neutral*, respectively:

the temperature of a lifted parcel remains similar to that of the (dry or moist) environment. Consequently, the parcel stays at the lifted position without vertical acceleration by the buoyancy force in such a situation.

2.1.3 Condensation levels and parcel theory

The Convective Available Potential Energy (CAPE) represents the hypothetical energy an air parcel would possess if it was pseudo-adiabatically lifted from the Level of Free Convection (LFC) to the Equilibrium Level (EL). Essentially, CAPE represents the energy readily accessible for convection and signifies the potential of an air parcel to gain buoyancy (Moncrieff and Miller, 1976). CAPE can be calculated as:

$$CAPE = g \int_{LFC}^{EL} \frac{T_v - \overline{T_v}}{\overline{T_v}} dz \tag{2.21}$$

with T_v for the virtual temperature of the air parcel and $\overline{T_v}$ for the virtual temperature of environment.

Meanwhile, when an air parcel is accelerated in an conditionally unstable environment as described in the previous section 2.1.2, it must overcome a certain amount of Convective Inhibition (CIN) and rise to a certain height where condensation begins. CIN signifies the energy quantity that would hinder an air parcel from ascending from the surface up to the Level of Free Convection (LFC). In other words, CIN quantifies the energy needed to counteract the negative buoyant force exerted by the surrounding environment on the air parcel.

A so-called Skew T - log p diagram (Figure 2.1) can give an illustrative overview of the identification of the different important levels listed below.

Lifting condensation level:

When an air parcel is lifted by external forces, for example, by orographic lifting over a mountainous barrier or by secondary circulation at a front, the height at which condensation commences is referred to as the Lifting Condensation Level (LCL). In case of forced ascent, the LCL represents the cloud base height. Using the Skew T - log p chart, the LCL is identifiable as the point where the dry adiabatic curve (which originates from the temperature close to the surface) intersects with the line representing a constant saturation mixing ratio (determined by the surface dew point or mixed over a certain lower layer).

Level of free convection:

In a conditionally unstable atmospheric stratification, an air parcel may be externally forced to ascend to a specific altitude. At this altitude, known as the Level of Free Convection (LFC), the parcel becomes warmer than its surrounding environment and subsequently undergoes positive acceleration by buoyancy force. This LFC can be pinpointed in a Skew T - log p diagram as the intersection point between the moist adiabatic curve originating from the Lifting Condensation Level (LCL) and the vertical temperature profile.

Equilibrium level:

The Equilibrium Level (EL) is virtually located at the second intersection where the air parcel's temperature matches temperature of its environment, causing its upward acceleration to cease.
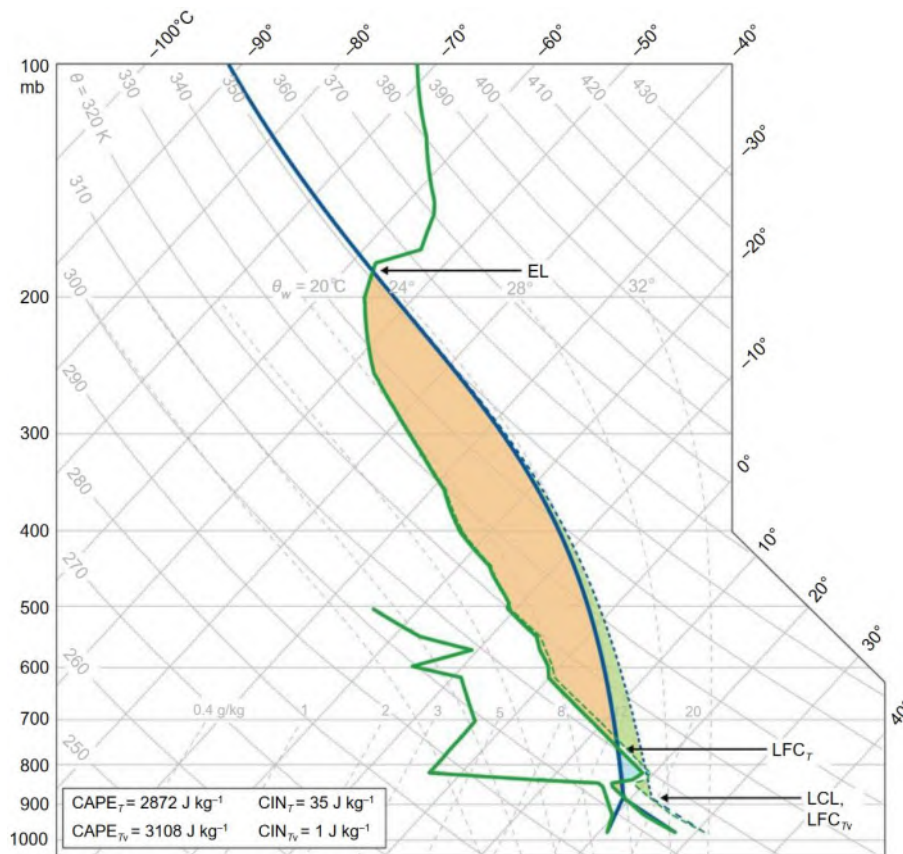
Figure 2.1: Skew T - log p diagram obtained from the Del Rio, Texas, sounding at 1800 UTC 14 May 2008. The profiles for temperature and dewpoint are represented by a solid green line, with dewpoints ceasing to be plotted above 500 hPa due to their unreliability at such heights. The virtual temperature profile is marked by a dashed green line. A solid blue line illustrates the temperature of an air parcel being lifted from the surface to its LCL, LFC, and EL. The virtual temperature of the air parcel is represented by a dashed blue line. Light orange and light blue shaded areas correspond to positive and negative areas of the solid blue parcel trajectory, respectively. These areas are proportional to $CAPE_T$ and CIN_T. The positive area of the dashed blue parcel trajectory, proportional to $CAPE_{Tv}$ (while the negative area is virtually nonexistent), extends over both the light orange and the light green shaded areas. (originated from Figure 2.9 in textbook of Markowski and Richardson (2011)).

The EL can be identified on a chart where the moist adiabatic curve and LFC, intersects with the temperature curve. In cases of deep moist convection, this level often aligns with the tropopause height. Due to the inertia of air parcels, the actual EL might extend a few hundred meters above the EL, particularly when clouds penetrate into the tropopause, a distinctive trait of intense cumulonimbus clouds.

In parcel theory, an air parcel encounters 3 thermodynamic stages, while it is lifted and rises through the troposphere:

1) **Dry adiabatic process**: the moist unsaturated air parcel from the ground surface rises along the dry adiabatic lapse rate until it becomes saturated and reaches the LCL, where condensation begins (Lamb and Verlinde, 2011). During this process, the potential temperature and mixing ratio within the air parcel keep constant.

2) **Moist adiabatic process**: after reaching the LCL, the lifted air parcel ascends along the moist adiabatic lapse rate. Latent heat is released through condensation of water vapor during further ascent, so that its temperature decreases at a slower rate.

3) **Buoyancy-driven process**: in case of positive buoyancy, and after reaching the LCL, the air parcel rises further along the moist adiabatic lapse rate. It is vertically accelerated by the positive buoyancy force driven by release of CAPE away from its initial position.

Thermodynamic velocity:

To quantify the vertical velocity of an air parcle during ascent, both sides of Eq. (2.11) are multiplied by $w = \mathrm{d}z/\mathrm{d}t$, leading to:

$$w\frac{\mathrm{d}w}{\mathrm{d}t} = B\frac{\mathrm{d}z}{\mathrm{d}t} \tag{2.22}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{w^2}{2}\right) = B\frac{\mathrm{d}z}{\mathrm{d}t} \tag{2.23}$$

Next, Eq.(2.22) is integrated over the time the air parcel requires to be lifted from the LFC to the EL; $w = 0$ at the LFC is also assumed as no buoyancy force is acting here. As a consequence of the ongoing buoyancy and, thus the vertical acceleration of the air parcel during ascent, the maximum vertical velocity w_{max} is assumed at the EL. The integration of Eq. (2.22) yields

$$\int_{EL}^{LFC} \mathrm{d}w^2 = 2\int_{LFC}^{EL} B\mathrm{d}z \tag{2.24}$$

$$w_{EL}^2 - w_{LFC}^2 = 2\int_{LFC}^{EL} B\mathrm{d}z \tag{2.25}$$

$$w_{max}^2 = 2\int_{LFC}^{EL} B\mathrm{d}z \tag{2.26}$$

$$w_{max} = \sqrt{2CAPE} \tag{2.27}$$

The value of w_{max} calculated by Eq. (2.26) can be taken as the upper limit of vertical velocity when the buoyancy force is only considered; sometimes w_{max} is also called the thermodynamic speed limit.

However, this approximation is typically overestimating the observed velocities in the real atmosphere owing to the absence of vertical perturbation pressures in the parcel theory, which plays an important role in the energy transfer. The parcel theory also neglects the exchange of momentum, moisture, and temperature between the parcel and its environment, referred to as entrainment (Markowski and Richardson, 2011). The mixing of environmental air into the air parcel during its rising leads to dilution and, thus, a slow down by decreasing the upward momentum and buoyancy.

2.2 Types of convective organizations

Thunderstorms are divided into four main types depending on their physical characteristics, formation, intensity and also its posed threat. Single-cells, and supercells are isolated convective systems, which are mostly commonly observed, whereas mesoscale convective systems (MCSs) as any ensemble of thunderstorms, which can produce hail, such as squall lines and mesoscale convetive complexes (MCCs), usually develop from isolated convection. They likely generate significant lightning flash rates and heavy precipitation. Hail events usually occur in multi- and supercells or mesoscale convective systems due to their large horizontal and vertical extension.

2.2.1 Single-cell convection

Single-cell convection is defined as a convective cell with only one single updraft without further subsequent initiation of any other cells. In a low-shear environment, a single cell thunderstorm can create a gust front, leading to shallow and weak updrafts. However, these conditions are not sufficient to trigger new cells. Its diameter is normally between 2 and 10 km, with a relative short longevity of about 30 minutes to 1 hour.

In an environment favorable for single-cell convection, a low vertical wind shear usually indicates weak synoptic-scale forcing. It is given that normally the diurnal cycle of the boundary layer plays an important role in the development of maintaining the single-cell convection rather than the large-scale ascent or the advection of temperature and moisture. As a result, single-cell convection is more likely to occur shortly after the maximum daytime heating is reached, and tends to dissipate quickly in the early evening. There is low possibility that single-cell convection produces significant hail or severe wind gusts (Markowski and Richardson, 2011).

The life cycle of this thunderstorm type constitutes of three stages (shown in Figure 2.2): towering cumulus, mature and dissipating stage. During the early stage of cumulus, only an updraft is present. Within an unstable environment, the air parcel is turning from unsaturated to saturated ascent when condensation begins (See Section 2.1.3). Cloud droplets start to grow and further increase their size. The mature stage is reached when the precipitation starts to fall from the cloud, based on the Bergeron-Findeisen process which explained the growth of ice crystal in the mixed phase cloud after reaching the freezing level. Precipitation particles, which also reduce the updraft, grow to sufficiently large size and fall down to the ground. At this time, a cool downdraft is generated in front of the precipitation area along the direction of movement with joint action of falling precipitation and subsequent evaporation. It eventually comes to the dissipating stage until the convection is entirely controlled by the downdraft. The updraft is cut off from the environment due to the cold outflow, which is more or less symmetrical owing to the low vertical wind shear.
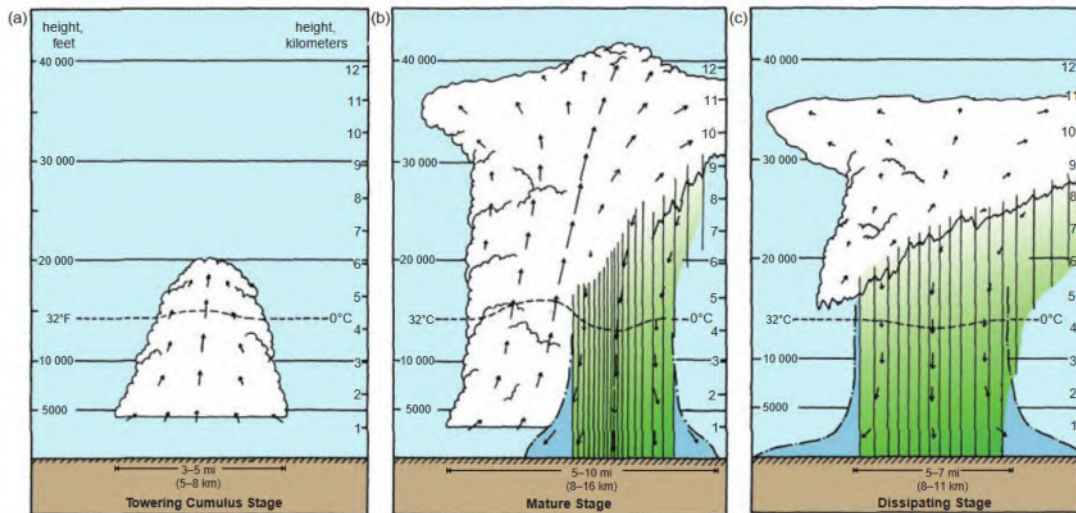
Figure 2.2: Three stages of a single cell: (a) towering cumulus stage, (b) mature stage, and (c) dissipating stage; (From Markowski and Richardson (2011), based on Byers and Braham (1949) and Doswell (1985)).

2.2.2 Multicellular convection

What makes multicell convective storms different from the single-cell is a relatively large and sufficiently strong gust front at one preferred flank that lifts air parcels to the LFC. Therefore new cells can be initiated and developed repeatedly to maintain the development of the multicell system. Multicells can persist between 30 minutes up to several hours, among which the most severe multicells have the ability to produce damaging straight-line winds and hail with sizes of up to several centimeters.

The preverred environment of multicellular convection is a medium degree of 0-6 km wind shear difference vector in the range between 10 - 20 ms^{-1} , which is higher than in typical single-cell convection environments (Markowski and Richardson, 2011). The gust-front lifting can lead to the propagation of the multicell system. A characteristic evolution of multicellular convection is shown in the Figure 2.3. Cell 1 is transiting into the dissipating stage, while cell 2 is currently in its mature stage. At the same time, cell 3 is just in the stage where the precipitation starts to take place, thus towards to the mature stage. The triggered new cell, cell 4, is forming usually at the right flank of the cluster and strongly influenced the by presence of horizontal vorticity induced by vertical wind shear and vorticity due to the cold pool (Marwitz, 1972; Weisman and Klemp, 1982). Every initiated new cell experiences the life cycle of a single-cell. They follow this life cycle one-by-one, e.g., Cell 3 in its early cumulus stage on the top panel into dissipating stage in the bottom panel.

2.2.3 Supercellular convection

Supercells are less commonly observed worldwide but burden on a large fraction for the responsibility of the damage from natural disasters. They can generate the hail with the large diameter due to the relative long-lived lifetime from 4 to 8 hours roughly. The widely accepted standard to
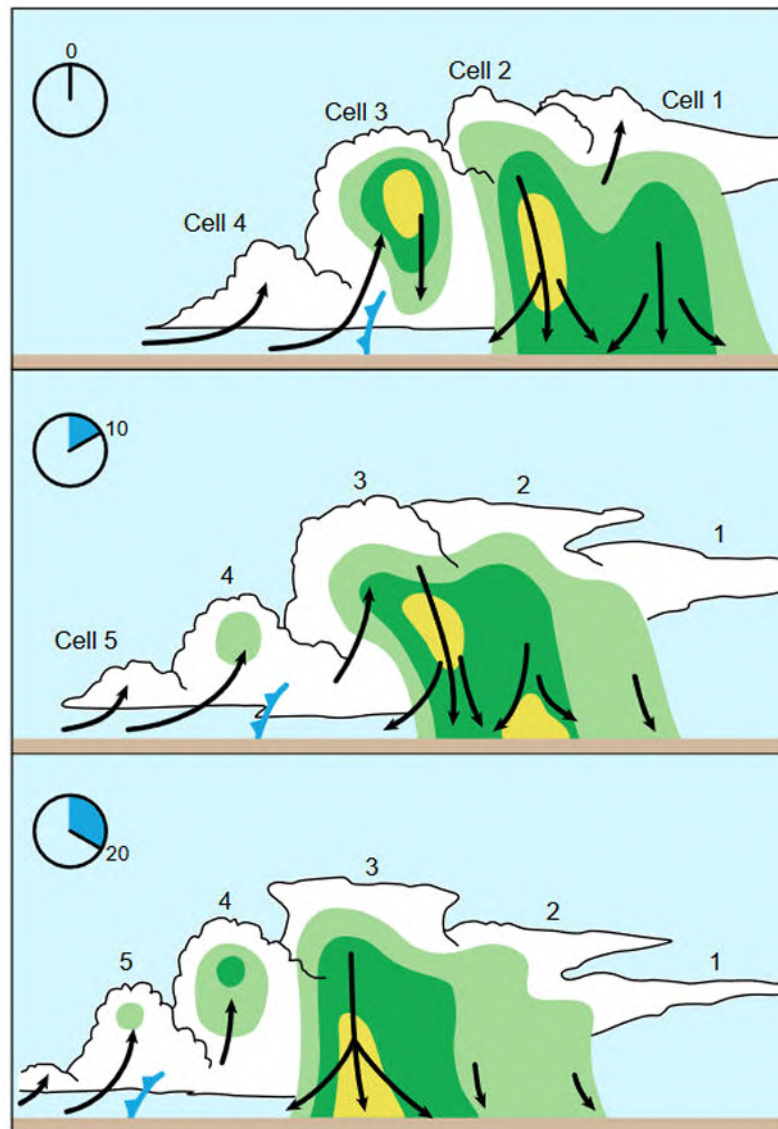
Figure 2.3: Temporal evolution of multi-cells. Colored areas are radar reflectivity, in which more centered location of area means stronger intensity; the initiated convection cells are numbered from 1-5; from the top to bottom are scenarios changing with time. More description for the process from Markowski and Richardson (2011), which is adapted from Doswell (1985).

identify a supercell is the presence of a long-lasting, deep mesocyclone within a strong updraft, with a mesocyclone width in the range of 3-8 km (Markowski and Richardson, 2011). Mesocyclones receive vertical vorticity from horizontal vorticity caused by large vertical wind shear, as a result of the vorticity tilting.

Supercells contain two regions of downdraft (See Figure 2.4). One is located to the rear of the storm, related to the hook echo region, which can often be identified in radar reflectivity data, indicating the strong updraft region within the supercell. This downdraft is denoted to as *rear-flank downdraft* or *RFD*. One possible explanation is that RFDs form when mid- and upper-level dry winds impinge upon the backside of an updraft, resulting in evaporative cooling and negative buoyancy to further enhance the downdraft. Another explanation is that dynamical vertical pressure gradient forces on the upshear flank of supercells could also trigger RFDs in the downward direction. However,
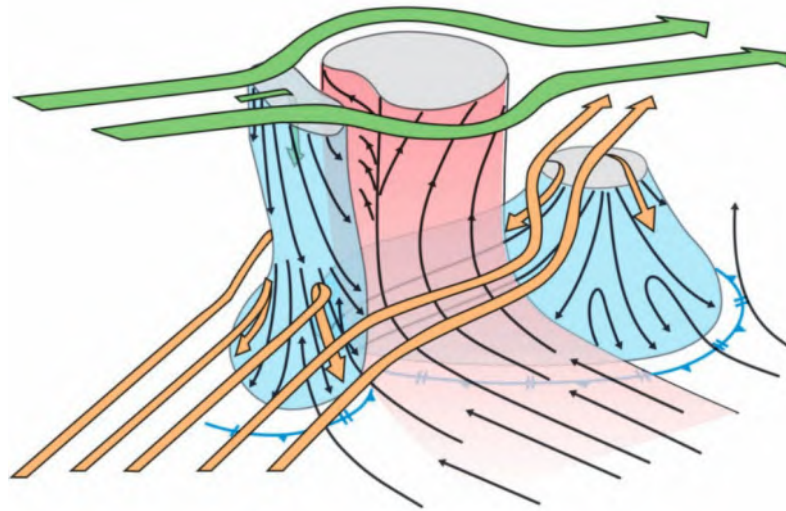
Figure 2.4: Schematic 3-D structure of the supercell. Black arrows is the streamlines. Blue-colored area on the left (right) side is where RFD (FFD) located. Red-colored area represents the updraft region. Green arrows represent 9 km flow while yellow arrows represent 4 km flow. (From Markowski and Richardson (2011), adapted from Lemon and Doswell III (1979)).

how strong the contributions from thermodynamic and dynamic forcings are in reality is still debated. Another downdraft associated with the sublimation of ice and evaporation of rain is named *forward-flank downdraft* or *FFD* and also heavily influenced by the low-level humidity and the entrainment of environmental air. *FFD* is usually more important for the supercell dynamics.

2.2.4 Mesoscale convective systems

Mesocale convective systems (MCSs) define a group of thunderstorms that create a continuous precipitation area with a spatial extent of at least 100 km on one axis (Markowski and Richardson, 2011). Mesoscale convective complexes (MCCs) as a sub-class of MCSs which are characterized by the tepmerature and spatial extent of their anvils and the larger circular cold Cumulonimbus (Maddox et al., 1986). Broad middle- and upper-tropospheric trough is a characteristic of synoptic weather patterns that favor MCCs (Markowski and Richardson, 2011). Squall lines as another sub-class of MCSs are a group of thunderstorms arranged in a line typically move swiftly, producing the heavy precipitation. Their length can extend for hundreds of miles, although their width is usually confined to about 10 to 20 miles (Markowski and Richardson, 2011).

Owing to their relatively large size and long duration, MCSs can produce a meso-α-scale (range from 200 to 1000 km) circulation with a warm-core cyclonic eddy in form of a convergent inflow below the level of convective heating, the so-called *mesoscale convective vortex*. Meanwhile, an anticyclonic mesohigh is formed near the tropopause with the divergent outflow under the impact of Coriolis acceleration. Figure 2.5 displays the configuration of different layers of MCCs as an example.
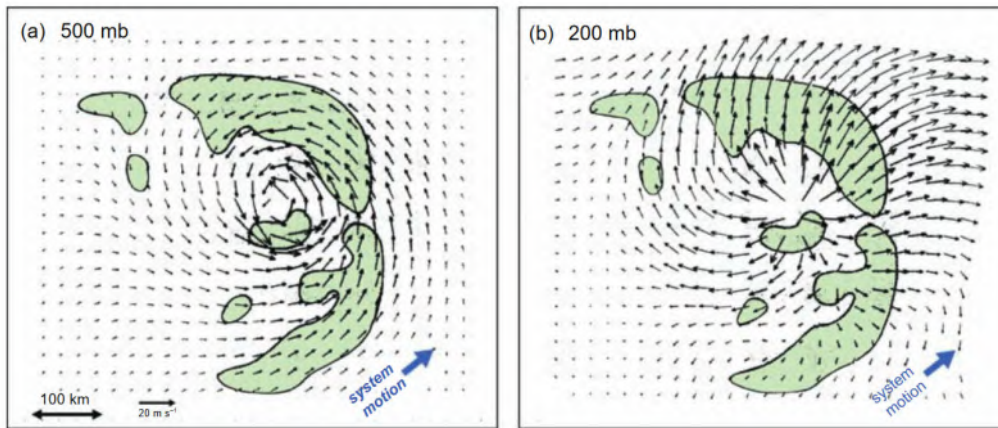
Figure 2.5: Vorticity distribution of MCCs from model simulation at 500 hPa (a) and 200 hPa (b), the shaded denotes precipitation area. (From Markowski and Richardson (2011), modified from Chen and Frank (1993)).

2.3 Development of hail

The size increase of hailstones is influenced by the complex processes between the airflow inside a deep convective cloud and precipitation particles. Hailstones primarily grow by accumulating supercooled cloud droplets and raindrops. A strong updraft within a thunderstorm is crucial for the development of hailstones, and deep convective storms have been observed to generate hail with the updraft speeds of up to 40 m s^{-1} (Crook, 1996; Maddox et al., 1986). The updraft carries the cloud droplets and small raindrops to the cold regions within the thunderstorms. Meanwhile, these supercooled hydrometeors do not immediately freeze, but stay in the liquid phase until the temperature of less $- 20$ °C, because of the limitation in the number of the ice nuclei particles. If supercooled droplets collide with ice particles moving from the rear downdraft region to the updraft region, the drops may freeze upon the surface of ice particles. This process causes the ice particles to grow to a size of one to five millimeters, which are called graupel particles (Markowski and Richardson, 2011). In this type of growth, growth is referred to as riming or collection growth and increases dramatically as drop size increases (Pruppacher et al., 1998).

As the hailstone goes on with collection growth, its fall velocity continues to increase, and the rate of growth also keeps increasing as long as the hailstone remains in a region where supercooled liquid water content is large (Markowski and Richardson, 2011). Their fall velocity is dependent on their diameter from the empirical derivation (Pruppacher et al., 1998) and defined as:

$$v_t \approx 9D^{0.8}, \tag{2.28}$$

where D is the diameter in cm and v in m s^{-1}.

The formation of hail can be through 2 different types: in the process of dry growth, the hailstone's initial form, or embryo, does not reach or exceed the freezing/melting point, despite the latent heat released by water droplets freezing on the surface of the graupel particles. As a result of this

process, the hailstone develops an opaque structure due to the incorporation of air pockets within its growing layer (Pruppacher et al., 1998).

The other type refers to the wet growth. If the surface temperature of the hailstones increases above the melting point due to the latent heat releasing, the liquid water can flow into the air pockets, leading to the structure of hailstones becomes transparent. Due to the continuous change of different types of the formation, hailstones often have a layered structure within the convective storms. Eventually, hailstones fall down after leaving the updraft at the front of thunderstorms, due to their larger gravity compared to the drag force from the updraft (Markowski and Richardson, 2011).

2.4 Basics of radar measurements

Weather radars offer plenty of principal advantages compared to other atmospheric observation methods, for example, the ability to monitor the vertical structure of thunderstorms, and relatively high resolution in time and space. This section will first introduce the components of a meteorological radar and then discuss how radar data is applied for the detection of hail events. The difficulties and possible limitations causing the inaccuracy of hail detection are summarized in the last part of this section.

2.4.1 Components of meteorological radar

The development and use of radars began before and during World War II (Marshall et al., 1955; Brown, 1999). Radar stands for "Radio Detection And Ranging". A radar emits electromagnetic waves to detect targets, such as aircraft, spacecraft, terrain or hydrometeors. The components of a radar system are shown in Figure 2.6. They mainly consist of a transmitter, antenna, and receiver (Skolnik, 1962).

The modulated short-pulsed radar signal is generated by the transmitter and radiated into the space by the antenna until it detects the target. The received backscattering of targets, which is a small portion of the emitted radar signal (Hughes, 1983), is caught by the antenna. This means that a duplexer allows a single antenna transmitting and receiving the signal in time-parallel. The returned echo is then collected and amplified by the receiver. The processed signal from backward scattering is transmitted and calibrated specifically for target detection, which is then visualized on the display monitor. This operation forms the core basis of radar functionality. A radar detect not only the signal strength, but also the range and angle of a target. Moreover, the output of radar can also provide information about the nature of a target.

How different the applied radar bands depend on the wavelength and the shape of the target. For meteorological purposes, k-, X-, C-, S-, and L-band radars are basically can be used for the detection of hydrometeors, even though the most common scanning radars operationally applied are C- and S-band radars. The ranges of frequency and wavelength for different meteorological radar bands
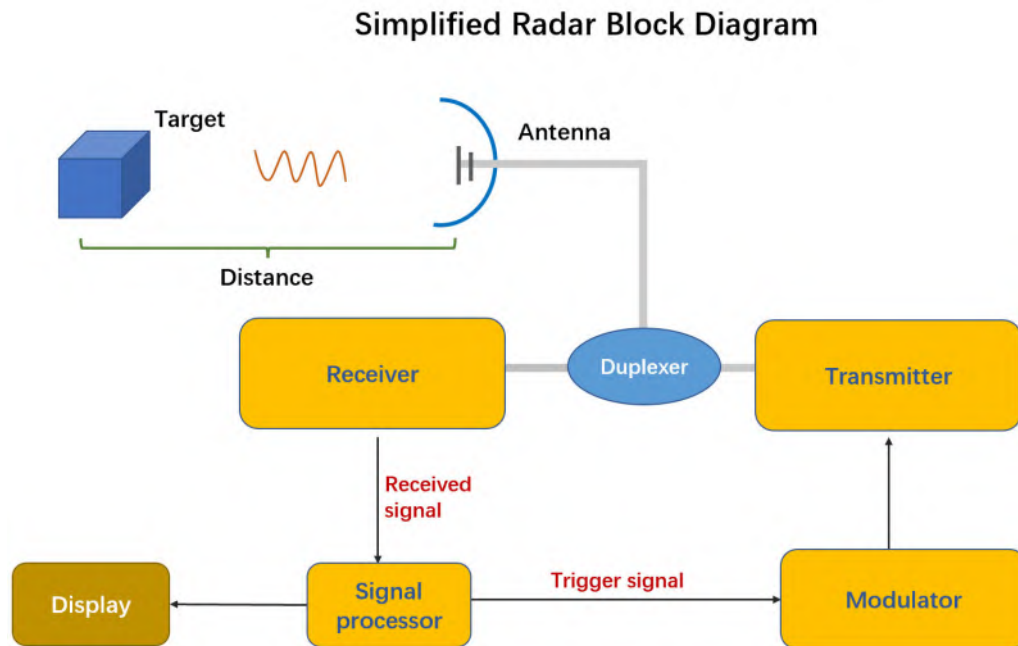
Simplified Radar Block Diagram

Figure 2.6: Block diagram of a meteorological radar. Modified from Fluck (2018) and Zrnić et al. (2014).

are summarized in Table 2.1. Bands with higher frequency such as K-, Ka-, Ku-, V- or W-band radar are heavily attenuated by atmospheric gases, cloud particles, or precipitation, therefore are usually used for other special detection of targets, e.g., for cloud observations.

Radar band	Frequency(GHz)	Wavelength(cm)
X	8 - 12.5	3.75 – 2.4
C	4 - 8	7.5 - 3.75
S	2 - 4	15 – 7.5
L	1 - 2	30 - 15

Table 2.1: Summary of different radar bands (Parker, 2010)

2.4.2 Radar detection of hail

In the short and medium distance of radar observations, the attenuation effect during the absorption and backscattering significantly influence the characteristics of the signal received from various targets. For meteorological observations, targets are usually hydrometeors both in liquid and solid forms. Radar reflectivity η is the characteristic quantity of the target at standard wavelength, but it is not able to represent the nature of the scattering medium (Sauvageot, 1992).

The radar reflectivity is defined as:

$$\eta = \int_{D_{min}}^{D_{max}} \sigma(D)N(D)dD \tag{2.29}$$

where $N(D)$ is the particle size distribution, and $\sigma(D)$ is the backscattering cross section with diameter D of the targets.

For the quantification of the radar reflectivity, the so-called Rayleigh approximation is used when the size parameter of the backscattering particles, $\alpha = \pi D \lambda$, is much smaller than the wavelength λ of the radar. This mainly depends on the wavelength (Sauvageot, 1992). The effective radar reflectivity of a spherical particle is approximately calculated by an equivalent dipole with the same diameter of the sphere. Then the cross section for the Rayleigh approximation can be simplified as:

$$\sigma = \frac{\pi^5}{\lambda^4} |K|^2 D^6 \tag{2.30}$$

where $K = |\frac{m^2-1}{m^2+2}|^2$, and m represents the complex refractive index, which is related to the dielectric permittivity and absorption coefficient of the scattering material. The actual value of $|K|^2$ is also influenced by the temperature and the wavelength of radar. Gunn and East (1954) presented corresponding calculations: For liquid water, $|K|^2$ is approximately 0.93, whereas for ice with normal density ($\rho = 0.92 \text{g cm}^{-3}$), $|K|^2 \approx 0.176$, independent of the temperature (Sauvageot, 1992). Eq. (2.30) is named as Rayleigh Law and is the basis for the application of microwaves to observe clouds and precipitation. It is assumed that the backscattering of the particles is isotropic.

By inserting Eq.(2.30) into Eq.(2.29), we can derive the formula for the radar reflectivity under the assumption of Rayleigh approximation:

$$\eta = \frac{\pi^5}{\lambda^4} |K|^2 \int_{D_{min}}^{D_{max}} D^6 N(D) dD. \tag{2.31}$$

The integral is the radar reflectivity factor Z:

$$Z = \int_{D_{min}}^{D_{max}} D^6 N(D) dD . \tag{2.32}$$

The radar reflectivity factor Z is an average distribution of scattering targets in the pulse volume and independent on the wavelength. This makes it easier to compare the outputs gained with different radars (Sauvageot, 1992). Especially for hail detection, determining the true value of Z is facing larger uncertainties, as the sixth power of the inhomogeneous diameter of hail is usually replaced by an equivalent diameter of spherical liquid droplets for the measurement.

The different units between Z in $m^6 \text{ m}^{-3}$ and size of precipitation in $mm^6 \text{ m}^{-3}$ can be simply transferred and the radar reflectivity is usually quantified in logarithmic units:

$$Z(dBZ) = 10 \ \log[Z(mm^6 \text{ m}^{-3})] \tag{2.33}$$

with the reference level $Z = 1 \ mm^6 \text{ m}^{-3} = 0 \ dBZ$.

As adumbrated above, the scattering also depends on the shape of the hydrometeors. Only small liquid droplets in the atmosphere are spherical. The other forms of hydrometeors are either nonspherical or near-spherical. For hail, as considered in this study, the dielectric factor also needs

to be taken into account when the water layer grows with relatively positive surface temperature and turns into transparent and dense ice progressively due to the release of latent heat. More importantly, a more accurate scattering approximation of hail is supposed to use the cross section of Mie scattering derived from the solutions of Maxwell's equations (Chỳlek, 1976), owing to the larger diameter of the scattering body compared to the wavelength. But Rayleigh approximation is still used when detecting hail as a result of the missing information about hydrometeor sizes to consider Mie scattering during the radar measurements (Jones, 1979).

Detecting hail by radar has always been facing difficulties in the past. Byers and Braham (1949) and other scientists tried to detect hail for the first time to study the properties and structure of a thunderstorm associated with the formation of hail. The complex relationships among the scattering effects of particles, horizontal extent of hail tracks and surface conditions studied by (Changnon Jr, 1970) rose more confusion regarding this question. Recent studies have attempted to estimate hail from radar by applying different approaches using both two- and three-dimensional (2D, 3D) radar data (Kunz and Kugel, 2015; Fluck, 2018; Schmidberger, 2018), but eventually have not established correlations between radar and hail characteristics.

The simplest criterion for assessing the presence of hail in a thunderstorm is quantifying the vertical maximum value of the radar reflectivity factor Z_{max}. This is referred to as Mason's criterion (Mason, 1971), when the radar reflectivity factor is

$$Z \geq 55 \text{ dBZ} .\tag{2.34}$$

This criterion suggest that hail is occurring also at the surface. However, owing to the limitations of radars as discussed below, the Mason criterion is only an estimate, and therefore a proxy of hail. Also very large raindrops can also reach to this threshold. Still, it is a valuable approximation for hail detection in plenty of studies (Kunz et al., 2020; Puskeiler et al., 2016; Holleman et al., 2000).

2.4.3 Inaccuracies in radar and limitations

Ground clutter

Radar echoes from ground clutter mixed with the echoes of meteorological targets can result in inaccuracies in radar data. This is particularly an issue for the hydrometeorological measurement (precipitation scan) at low elevations, where objects like chains of hills and also high buildings in the city or wind turbines are detected (Billingsley, 1993). This problem can be fixed by a **Doppler** filter with zero frequency (Sauvageot, 1992).

Attenuation

Attenuation describes the signal loss by absorption and scattering between radar and targets by hydrometeors and gases of the air.

Mainly oxygen and water vapor lead to larger attenuation effects (Van Vleck, 1947a,b), which can be described as a function of frequency, and their attentuation remains below 10 GHz (Bean and Dutton, 1966). For hydrometeors, attenuation from cloud and precipitation is discussed separately. The attenuation of cloud particles with adequately small radioelectric dimensions reach the criterion of Rayleigh approximation. The attenuation in clouds increases with increasing temperature and when wavelength decreases (Gunn and East, 1954). The attenuation by rain, snow and hail is determined differently due to different particle sizes. Especially attenuation by hail with large particles, due to its dielectric properties, non-spherical shape and unstable surface condition, can not be calculated unless all of the physical properties can be sucifficiently obtained, which is usually not the case (Sauvageot, 1992).

Beam shielding and other effects

The radar signal for regions behind mountains can be shielded by the obstacles, leading to lower or gap composites in the reflectivity of the targets. This shielding effect can even distort certain captured targets in specific situations.

In radar systems, nonuniform beam filling with hydrometeors can cause a reduction in the core reflectivity. This is the case especially for relatively large beamwidth and pulse lengths, along with power distribution across the beam. This nonuniformity can adversely affect the measurement of precipitation such as rain or hail, because only a small portion of the total pulse volume might be impacted by the targets (Atlas, 2015). At the same time, wide beam scanning can decrease the spatial resolution to identify convective cells at very long distance.

2.5 Machine learning algorithm

In contrast to traditional numerical weather prediction (NWP) modes, which are to solve the governing equations by numerical methods, machine learning (ML) models are not directly programmed but trained Chollet (2021). The models are presented with a massive number of relevant examples related to the prediction task, and they find the statistical relationships which allow the system to automatically come up with rules to tackle the task. The development of artificial intelligence (AI) and ML can also benefit meteorological studies as an advanced tool to reveal the physical mechanism behind certain phenomena. ML has already been widely used in many directions such as prediction, classification and identification, and shown its potential worth in current atmospheric researches.

Neural networks were successfully applied, for example, to predict El Niño related to climate change (Hsieh and Tang, 1998), and also for climate attribution (Pasini et al., 2017). Since the convolutional neural networks was developed by Krizhevsky et al. (2012), they become soon implemented into different studies, for example, to analyze the spatial feature of severe hailstorm (Gagne II et al., 2019). An example from Scher and Messori (2018) using convolutional neural networks in order to estimate future forecast uncertainty from past forecasts displayed the improvement skill with lower

computational costs compared with ensemble forecasts. More information on the convolutional neural network used in this master thesis will be detailed in Section 3.2.5.

2.5.1 Three branches of Machine learning

ML can be separated into the following three learning strategies (Chollet, 2021):

Supervised learning:

It is the most common ML approach when finding the rules/relationships between the input data and the explicit target, such as optical character recognition, speech recognition, image classification, and language translation. It is also mainly used in this master's thesis.

Unsupervised learning:

This branch consists of grasping transformation from the input data without the help of the labelled target. Usually, it is used for data visualisation before supervised learning to better understand the correlation within the present dataset. For example, clustering is basically a collection of objects on the basis of similarity and dissimilarity between them as unsupervised learning.

Reinforcement learning:

Reinforcement learning refers to an agent added which can choose actions that maximize rewards based on information obtained from its environment so as that the learning can automatically continue towards the expected direction. Some more examples of reinforcement learning in image processing include, for example, Robots equipped with visual sensors from to learn their surrounding environment.

2.5.2 Overfitting and underfitting

The key point for the ML is to find the best position between optimization and generalization. Optimization refers to the procedure of modifying a model until reaching its best performance on the training data. Generalization reflects how well the trained model can perform on the untouched new data. *Underfitting*, indicating a bad optimization of the model, means that the model can not present a good performance neither on training data nor on the validation data. *Overfitting* refers to the fact that the model achieves better performance on the training data than on the validation data (Chollet, 2021). It happens when some noise in the training data is grasped up, which are treated incorrectly as signals by the model, meaning that generalization of the model can not meet our expectation.

In practice, overfitting is more common when training a model, and there are four common ways to prevent the overfitting: 1) Get more training data; 2) Reduce the capacity of the network; 3) Add weight regularization (Section 3.2.4); 4) Add dropout (Section 3.2.5).

3 Data and Methods

As mentioned in section 1, owing to the limitation of direct ground observation data for hail, this study uses proxy data from two different data sets. The data set for the potential hail tracks is estimated from 3D radar data from DWD, using the specific storm-track algorithm TRACE3D (Handwerker, 2002; Schmidberger, 2018). The second data set characterizing ambient conditions of the thunderstorms is the ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF). Here, only the time period from April to September (summer-half year), when most of the damaging hail events occurs with high frequency (Kunz and Puskeiler, 2010; Punge and Kunz, 2016), is considered for the period 2005-2019. More information on both data sets and the applied algorithm and methods are shown in the following subsections in this chapter.

3.1 Hail data and Reanalysis data

3.1.1 Radar data

The German radar composites for the period 2005 to 2019 used in this study were provided by DWD, which operates a network of 17 C-band radar systems in Germany. The DWD has operated a radar network since 1994, which initially included five radar stations. Over the years, the network expanded to a total of 16 radar stations by 2000. The radar units installed since 1994 are C-band Doppler radars known as "METEOR 360 AC" with a wavelength of 5.4 cm, according to Bartels et al. (2005). To enhance the time resolution of the volume scan and to aid in the development of new, more effective algorithms, a precipitation scan strategy with 150 km radius for 5-minute volume scanning was successfully introduced in late 2012 with frequency of 600 Hz (Seltmann et al., 2013). As can be seen in Figure 3.1, Germany is almost completely covered by the precipitation scan with a radial range of 150 km (Helmert et al., 2014). This was done alongside the process of changing the radar system. The radar data, with an (interpolated) range resolution of 250 meters, is utilized in various applications including those related to warning management and hydrological detections (Helmert et al., 2014). To ensure accurate scanning of the atmosphere without interference from nearby obstacles, the radar units are usually placed on hills or towers. From 2011 to 2015, the radar units were gradually replaced with dual-pole Doppler radars. In 2013, the DWD added the Memmingen station to the radar network, bringing the total number of operational polarimetric radar systems from 16 to 17.
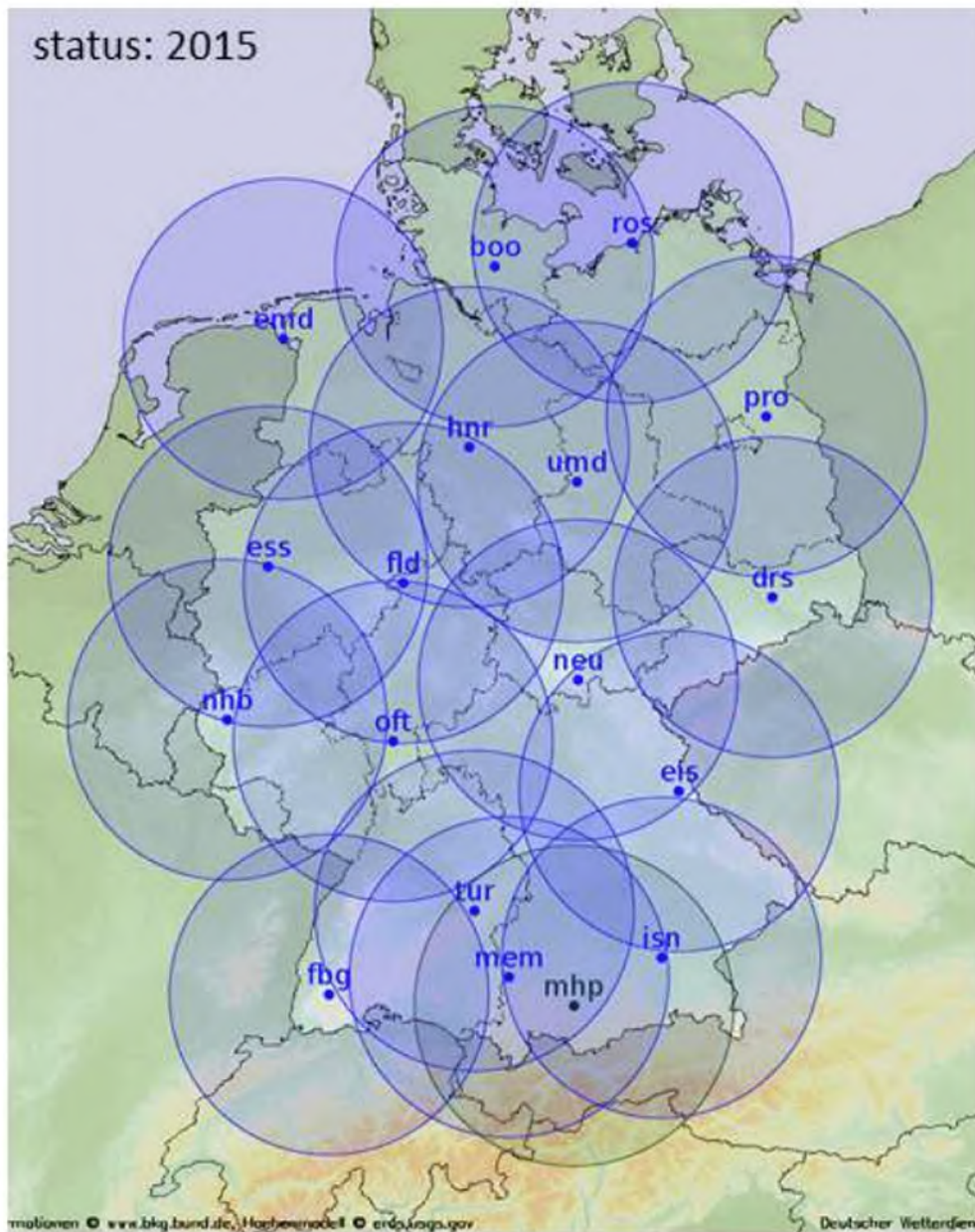
Figure 3.1: DWDs radar network on the basis of the precipitation scan with 150 km range after the introduction of polarimetric C-Band Doppler radar systems (From Helmert et al. (2014)). The names and location of radar stations are listed in Table 3.1.

Name of radar station	Abbreviation	WMO Nr.	Coordinate
Boostedt	BOO	10132	54°00′15.8″N
			10°02′48.7″E
Dresden	BRS	10488	51°07′28.7″N
			13°46′07.1″E
Eisberg	EIS	10780	49°32′26.4″N
			12°24′10.03″E
Emden	EMD	10204	53°20′19.4″N
			07°01′25.5″E
Essen	ESS	10410	51°24′20.2″N
			06°58′01.6″E
Feldberg	FBG	10908	47°52′25″N
			08°00′13″E
Flechtdorf	FLD	10440	51°18′40.31″N
			08°48′07.2″E
Hannover	HNR	10339	52°27′36.2″N
			09°41′40.2″E
Neuhaus	NEU	10557	50°30′00.4″N
			11°08′06.2″E
Neuheilenbach	NHB	10605	50°06′00.4″N
			06°32′53.9″E
Offenthal	OFT	10629	49°32′26.4″N
			12°24′10.03″E
Prötzel	PRO	10392	52°38′55.22″N
			13°51′29.57″E
Memmingen	MEM	10950	48°02′31.7″N
			10°13′09.2″E
Rostock	ROS	10169	54°10′32.4″N
			12°03′29.1″E
ISEN	ISN	10832	48°10′28.9″N
			12°06′06.3″E
Tükheim	TUR	10832	48°35′07″N
			09°46′58″E
Ummendorf	UMD	10356	52°09′36.3″N
			11°10′33.9″E

Table 3.1: Information of the radar network of the DWD

3.1.2 Storm-track algorithm TRACE3D

Many different algorithms for tracking SCSs have been developed in recent decades. Real-time automated tracking algorithms can manage to identify convective storms by defining the storms as a continuous region with high reflectivity in every scan of weather radar (e.g., Dixon and Wiener, 1993). TRACE3D, originally developed at IMK-TRO by Handwerker (2002) for a single C-band radar in spherical coordinates, includes two main steps: i) identification, and ii) tracking of reflectivity cores (RCs).

Almost all convective cells, accompanied within the large regions of intense precipitation (ROIP), are first identified and extracted by using the tracking algorithm (Handwerker, 2002). In figure 3.2 it is shown how the first step of identification works. The presence of an ROIP is determined by the first threshold dBZ_{min} with $\geq 35\ dBZ$ of radar reflectivity. This threshold value guarantees that most of the thunderstorms are discovered and easily considered as a whole since the algorithm is not able to capture all individual features of every storm. Then the difference dBZ_{diff} ($\simeq 10\ dBZ$) is subtracted from their multiple relative maximum in every ROIP. Contiguous regions above the second threshold in each ROIP are adaptively determined ($dBZ_{max} \geq dBZ_{min} + dBZ_{diff} (= 45\ dBZ)$) as individual RCs. Note that RCs are only represented by the reflectivity values, but not by the storm volume.

Tracking is recorded only when RCs are identified sustainably both in the last previous scan and latest new scan with the condition of less than 30 minutes time lag in between. The parent RC in the earlier scan is moving towards a certain direction with its own velocity and location information at a known time point. Accordingly, the children RCs might show up and be captured in the later scan near the location of parent RC. Then the size and intensity of children RCs is compared with the parent's to characterize the possible evolution of the thunderstorm. Cell merging or cell splitting are thus also considered. Once merging or splitting happens, the further parents or children will be subsequently searched in the next scan.

Later, with the aim of investigating hailstorms for entire Germany, it was necessary to adapt the algorithm to the data of the German weather radar network from DWD. Three points form the major differences compared to the original use of cell tracking in a single radar in spherical coordinates (Schmidberger, 2018):

(i) Cell tracking for radar composite data,

(ii) Use of Cartesian coordinates (x, y, z) instead of the spherical coordinates (r, θ, φ),

(iii) Cell tracking with discrete reflectivity classes.

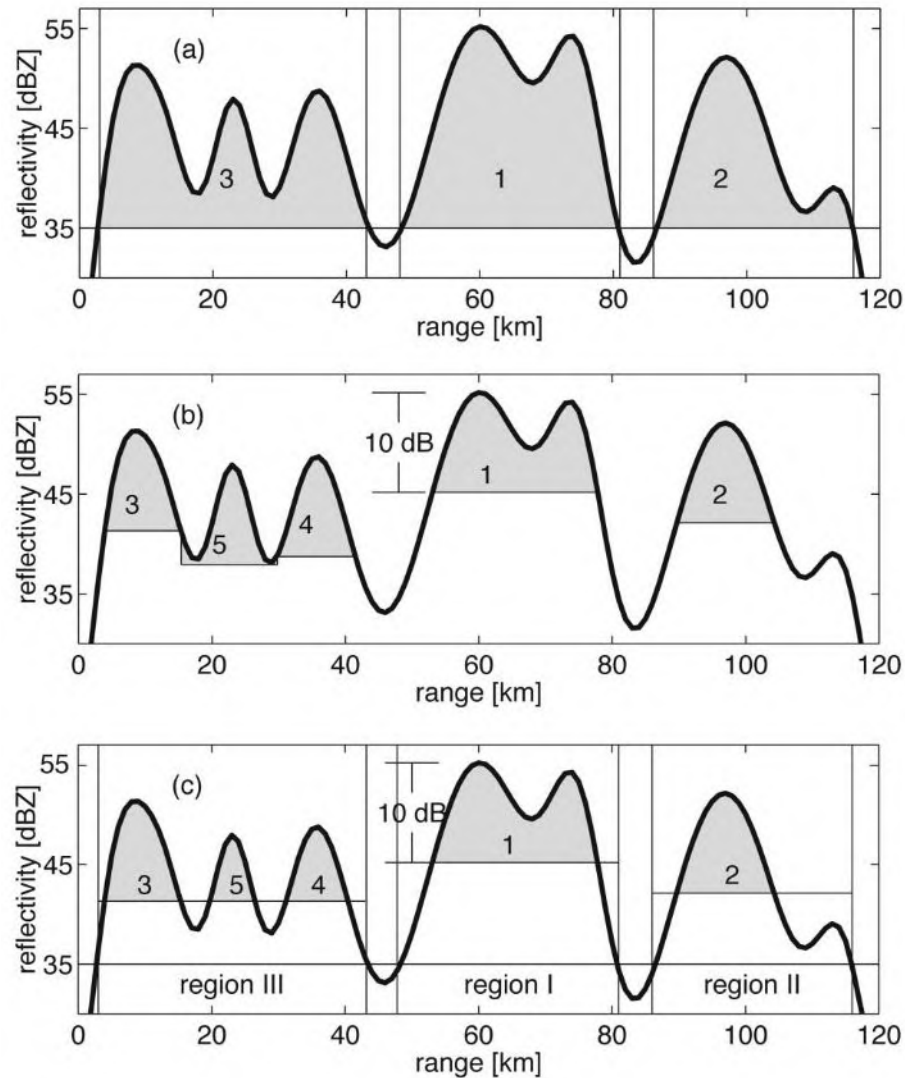Figure 3.3 shows the flowchart of the TRACE3D algorithm for the identification and tracking of RCs.

Figure 3.2: This figure depicts the method for defining reflectivity cores (RCs) in a one-dimensional radar data set. Reflectivity cores are defined as regions in which the radar reflectivity exceeds a certain threshold, known as dBZ_{min} (35 dBZ). a) considering all beam volume elements above the reflectivity threshold as part of an RC, b) selecting the upper 10 dB of reflectivity (dBZ_{diff}) as the RC, which may include lower regions of adjacent RCs, and c) searching for contiguous regions with reflectivity above the threshold and using a second threshold, 10 dBZ below the maximum reflectivity in each region and implemented by TRACE3D, is recommended as it effectively defines RCs while minimizing overlap between adjacent regions. These thresholds serve as a basis for identifying and characterizing reflectivity cores in radar, more details see (Handwerker, 2002).
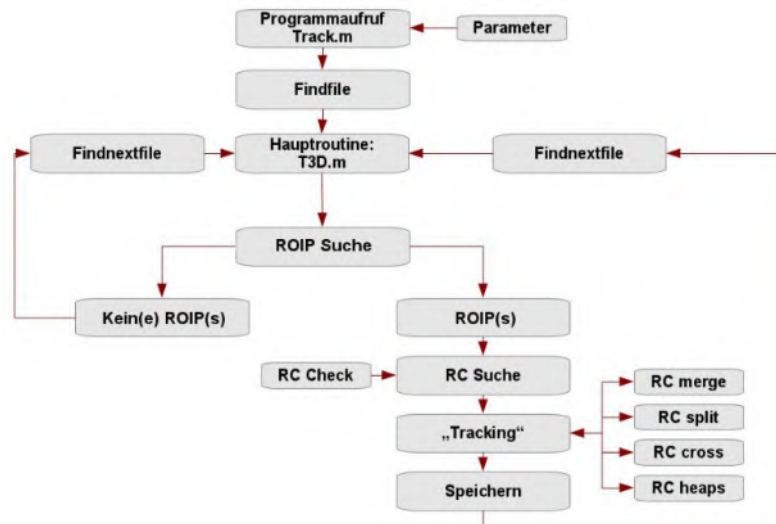
Figure 3.3: Flowchart illustrates the program structure and the interaction between the most important routines in TRACE3D (Figure from Schmidberger, 2018).

Detecting as many potential hail tracks as possible and extracting their pathway is the main goal. As to disregard uninterested weak convective events, S2 configuration (Schmidberger, 2018) is therefore set to serve the purpose.

S2: Detection with dBZ_{max} = 52 dBZ, dBZ_{diff} = 10 dBZ; RCs of class 59. 5 dBZ are detected here; adjacent reflectivities of class 50.5 dBZ are added to them.

According to Schmidberger (2018), configuration S2 results in a significant reduction in the number of cells and few cell divisions or cell fusions, while the S2 detects pathways of hailstorms already promisingly.

Prevailing errors in TRACE3D are the one-time crossing of the thresholds of nearby cells. The following corrections are implemented, with the aim to improve the accuracy of identifying hailstorm paths. Firstly, the incorrect linkages are removed depending on the minimum distance and pathway direction change criteria. Secondly, start point or end point of a pathway with errors is deleted. Last, faulty links leading to the trajectory offsets are excluded.

3.1.3 Hail-affected area

Based on the storm track algorithm in the Section 3.1.2, in total of 8184 estimated hail streaks is determined from 2005 to 2019 (Schmidberger, 2018). For each hail streak, its corresponding date time, estimated length and averaged width of streak, duration, direction of propagation, center location of the streaks and propagation velocity is extracted and recorded from the radar reflectivity of 3D radar in the Section 3.1.1.

Estimation of the hail-affected area is necessary and helpful to prevent hail damage and impacts in advance. While no storm detection and quantification technique can flawlessly capture every storm event, the implementation of the tracking algorithms (Section 3.1.2) and the function in this
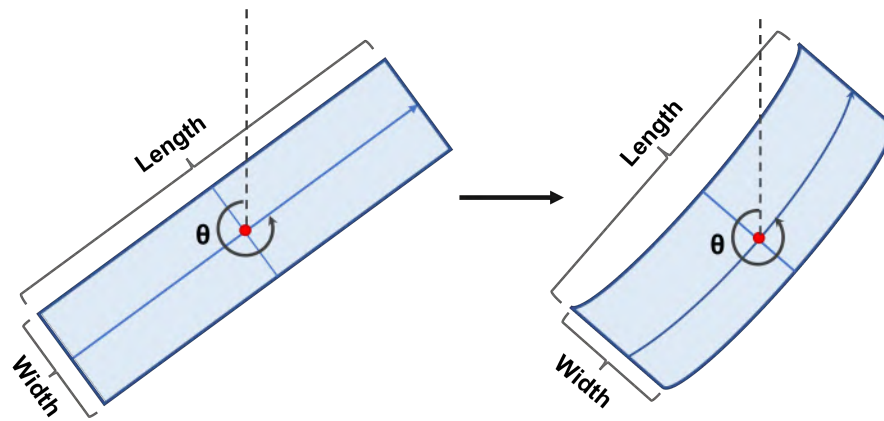
Figure 3.4: Demonstrative polygons of the area affected by a single hail streak, the left (right) is the area on a flat (curved Earth) surface, and the red dot is the centre of the hailstorm at a certain time, θ shows the direction of the hail track.

section generates a substantial data set of hailstorm occurrences for comprehensive assessment and evaluation. In this study, the hail-affected area for each hail streak is first estimated as a function of its length and width, derived from data of the TRACE3D storm track algorithm (see Section 3.1.2). The objective is to evaluate the potential area-averaged hail threat in terms of the spatial coverage at the local scale for the first time. Because of the large uncertainty included in both radar observations and the hail detection algorithm, it cannot be assured that each radar-identified track produced hail on the ground. Therefore we use here the term "potential" hail track.

The assessment of the hail-affected area is divided into two steps. According to Figure 3.4, the area affected by a single potential hail streak can be obtained by multiplying its length by its averaged width on a flat surface. At the same time, it can be further considered on the spherical curve of the Earth to implement the spatial hail estimation at the local scale. Although there remains some uncertainty due to the idealised features of hail streaks and their characteristics, it offers the only possibility to obtain a predictable reference value for the areas affected by hailstorms and to quantify the total coverage of potential hail threat. In this step, we focus only on daily aggregated affected areas for entire Germany during the hail-favorable season as target data for the ML method described later.

3.1.4 ERA-5 Reanalysis

Reanalysis data is a valuable tool for estimating the state of the atmosphere by combining the latest observations with a state-of-the-art NWP model. The process of Earth System Data Assimilation enables the creation of consistent and comprehensive data sets to understand weather patterns and climate change over time. Reanalyses can help to identify weather extremes, to assess forecast quality, and to understand climate change impacts by providing spatially complete and consistent data.

ERA-5 data sets is the latest replacement of ERA-Interim from ECMWF within the Copernicus Climate Change Service (C3S) and covers the period from 1950 to the present (Hersbach et al.,

2020; ECMWF, 2023). The assimilation system of the ERA-5 reanalysis utilizes the Integrated Forecasting System (IFS) Cy41r2 based on 4-dimensional variational analysis. Its spatial resolution extends to 31 km globally and 137 levels in the vertical up to 1 Pa. The data of the deterministic model is hourly recorded, and of the Ensemble of Data Assimilations (EDA) 3-hourly. Compared to ERA-Interim, ERA-5 contains uncertainty estimates from the 10-EDA, and more factors (e.g., sea-ice cover, volcanic eruption and greenhouse gas concentrations) are taken into count as model input (ECMWF, 2022). More updated details about ERA-5 is documented in Hersbach et al. (2020). In this study, we mainly focus on the horizontal environmental fields in Germany (Latitude: $44°N - 56°N$; Longtitude: $4°W - 16°E$) by analyzing grid point data with the horizontal resolution of $0.25° \times 0.25°$ (longtitude $\times$ latitude). Besides, many non-standard meteorological variables described in the next subsection 3.1.5 have been calculated at IMK-TRO from the raw reanalysis data.

Full name	Abbreviation	Unit
Zonal wind at 500 hPa	$U500$	m s^{-1}
Temperature at 850 hPa	$T850$	K
Temperature at 2 meter above the ground	$T2m$	K
Equivalent potential temperature at 850 hPa	$THETAE850$	K
Relative humidity at 850 hPa	$RH850$	%
Specific humidity at 850 hPaa	$Q850$	kg kg^{-1}
Wind speed 6km above ground	$VV6km$	m s^{-1}
Bulk wind shear	$BULKSH$	m s^{-1}
Lapse rate between 700 hPa and 500 hPa	$LAPSE_{700500}$	K km^{-1}
Convective available potential energy	$CAPE$	J kg^{-1}
Surface lifted index	SLI	K
Isothermal level of 0 degrees C.	$DEG0L$	m
Cape$\times$wind shear	$CAPESHEAR$	m^2 s^{-2}
Vertical totals index	VT	K
Total totals index	TT	K
Total column water	TCW	kg m^{-2}
Significant hail parameter	$SHIP$	-
Supercell composite Parameter	SCP	-

Table 3.2: Summary of environmental atmospheric parameters for ML.

3.1.5 Candidates of convective parameters

Numerous convective parameters, which play different roles in the convective mechanism or evolution of different types of convective storms, diversely reflect the storm environments. Some of the potential candidates covering different aspects are listed in Table 3.2. These convective

parameters when correlating with the potentially hail-affected area can also provide the reference for the ML model, to assist in identifying more favourable environments of hailstorms.

As background information for the ML model, the different atmospheric temperatures, humidity and general wind speed of the lower and middle troposphere are selected to set up and represent the large-scale atmosphere conditions (see Table 3.2).

Bulk wind shear

Bulk wind Shear is the difference in wind vectors over a layer of the atmosphere. When it is measured between 0-6 km above ground level (AGL), it can effectively distinguish between environments that are favorable for supercell thunderstorms and those that are not. The bulk wind shear can be also highly related to the deep layer shear (DLS), which favors the development of supercell thunderstorms (Wilhelm, 2022).

Lapse rate between 700 hPa and 500 hPa

The Lapse rate (defined in Section 2.1.1) provides information about the stability of an atmospheric layer. The faster the temperature decreases with height the more unstable is the atmosphere, namely decisive for parcel lifting (see Section 2.1.3) and the development of hailstorms. Here the lapse rate between 700 and 500 hPa is considered as the representation of the potential atmospheric instability (Kunz et al., 2020).

Convective available potential energy

CAPE describes the maximum amount of potential energy available for an air parcel for free convection; more details are given in the Sections 2.1.1 and 2.1.3. The values of CAPE can range from zero to over 5000 $J\ kg^{-}1$, and different ranges indicate varying degrees of atmospheric instability (Markowski and Richardson, 2011; Doswell III, 1987).

$$CAPE = \begin{cases} < 1000\ J/kg & \text{weak instability} \\ 1000 - 2500\ J/kg & \text{moderate instability} \\ 2500 - 4000\ J/kg & \text{strong instability} \\ > 4000\ J/kg & \text{extreme instability} \end{cases} \tag{3.1}$$

Surface lifted index

The surface lifted index (SLI) is an indicator of latent instability (Galway, 1956; Kunz et al., 2020). SLI is defined as the temperature difference between the environment and the parcel at 500 hPa (details in Section 2.1.3). The more negative the SLI values are, the more unstable are the atmospheric ambient conditions. Choosing SLI has the advantage for quantifying instability independent of the height of the EL (other than CAPE), which normally has a large bias from model output.

CAPE×wind shear

The CAPESHEAR is a measure that combines both the CAPE and the bulk wind shear (computed at ECMWF between 925 and 500 hPa). This parameter assesses the potential for well-organized and long-lived convective storms such as MCSs and supercells: CAPESHEAR >1500 $m^2\ s^{-2}$ suggests that there is potential for very active convection.

Vertical totals index

The Vertical Total (VT) is the temperature difference between the 850 and 500hPa levels and – like the lapse rate – a parameter of the average temperature gradient. It describes the temperature difference roughly between the top of the atmospheric boundary layer and the level halfway through the air mass. A stronger vertical temperature difference indicates a higher likelihood of thunderstorms. Typically, a threshold of 26 K is used to differentiate between weather conditions that are conducive to thunderstorm formation and those that are not. This value of VT is approximately equivalent to an average observed lapse rate of 0.65 K per 100 m (Grieser, 2012).

Total totals index

The Total Totals (TT) Index is a meteorological parameter that estimates the likelihood and intensity of thunderstorms by measuring the vertical temperature and humidity gradient. It calculates the temperature difference between the near-surface layer (850 hPa) and the mid-troposphere (500 hPa), similarly to the VT, and takes additionally into account the moisture content between these two layers. The larger TT is, the higher the likelihood for deep convection or the formation of thunderstorms (ECMWF, 2010b).

Total column water

Total column water (TCW) refers to the combined amount of water vapor, liquid water, cloud ice, rain, and snow present in a vertical column extending from the Earth's surface to the uppermost part of the atmosphere (ECMWF, 2010a). It can also quantitatively reflect the regional precipitation (e.g., hail, extreme rainfall) of thunderstorms and water vapor path during the development of thunderstorms.

Significant hail parameter

The Significant Hail Parameter (SHIP) is used to differentiate between hail with diameters of at least 2 cm (significant hail) and hail that is smaller (non-significant hail). SHIP values greater than 1 suggest that conditions are favorable for significant hail, with values greater than 4 considered extremely high. Typically, SHIP values larger than 1.5 to 2 are associated with observed instances of significant hail (Grieser, 2012; Prein and Holland, 2018) and is defined as below:

$$SHIP = \frac{1}{42\,000\,000} \frac{CAPE_{MU}}{J\,kg^{-1}} \frac{Q_{PARCEL,MU}}{g\,kg^{-1}} \frac{LAPSE_{700500}}{K\,km^{-1}} \frac{273.16\,K - T_{500hPa}}{K} \frac{BULKSH}{m\,s^{-1}}. \quad (3.2)$$

Supercell composite parameter

The definition of the supercell composite parameter (SCP) is based on 3 important factors: CAPE, Storm Relative Helicity (SRH) and DLS:

$$SCP = \frac{CAPE_{MU}}{1000 \; J \; kg^{-1}} \frac{SRH_{0-3km}}{100 \; m^2 \; s^{-2}} \frac{DLS}{20 \; m \; s^{-1}}, \tag{3.3}$$

where SRH is a measure of storm-relative streamwise vorticity used for supercell forecasting (Markowski and Richardson, 2011). Typically, when the SCP value is greater than 1, it indicates a high likelihood of supercell storms. Conversely, SCP values below 1 are usually associated with non-supercell storms. Thompson et al. (2003) conducted a study on this topic and reported that the average SCP value for supercell storms was 4, while it was 0.2 for non-supercell storms (it is refered to their study for more details).

3.2 Methods

3.2.1 Correlation analysis

One of the earliest and most widely used correlation coefficients is the Pearson correlation coefficient (Boddy and Smith, 2009), which measures the linear relationship between two variables. In meteorology, the Pearson coefficient has been used to study the relationship between various meteorological variables and their impacts on weather phenomena, including hailstorms and severe thunderstorms (Cohen et al., 2009; Zhang.Y and Niu.J, 2009). However, the Pearson correlation coefficient requires a normal distribution of the sample.

If this is not the case, the Spearman's coefficient should be used, which determines whether two non-linear relationships can be described by monotonic functions, without making any assumptions about the distribution of variables. Therefore, it is applied in this thesis to study relationships between ambient conditions of hailstorms and hail streakes. It is also suitable for variables when outliers are present (Hauke and Kossowski, 2011). The Spearman correlation coefficient ρ_s used in this thesis is defined by

$$\rho_s = \frac{\sum_i (Var_{env,i} - \overline{Var_{env,i}})(Hail_i - \overline{Hail_i})}{\sqrt{\sum_i (Var_{env,i} - \overline{Var_{env,i}})^2 \sum_i (Hail_i - \overline{Hail_i})^2}} \tag{3.4}$$

where i is the corresponding grid point; $\overline{Var_{env,i}}$ and $\overline{Hail_i}$ is the average of convective parameters and hail-related parameters respectively. $Var_{env,i}$ is the variables corresponding to ambient conditions of hailstorms and $Hail_i$ is the hail-related parameters (here the hail-affected area is chosen).

3.2.2 Correlation among the convective parameters

A large number of convection-relevant atmospheric environmental variables are used to improve the estimation of the expected life cycle characteristics of convective storms in the thesis of Wilhelm (2022). For each of the radar-derived storms recorded, a spatial average around the storms' location
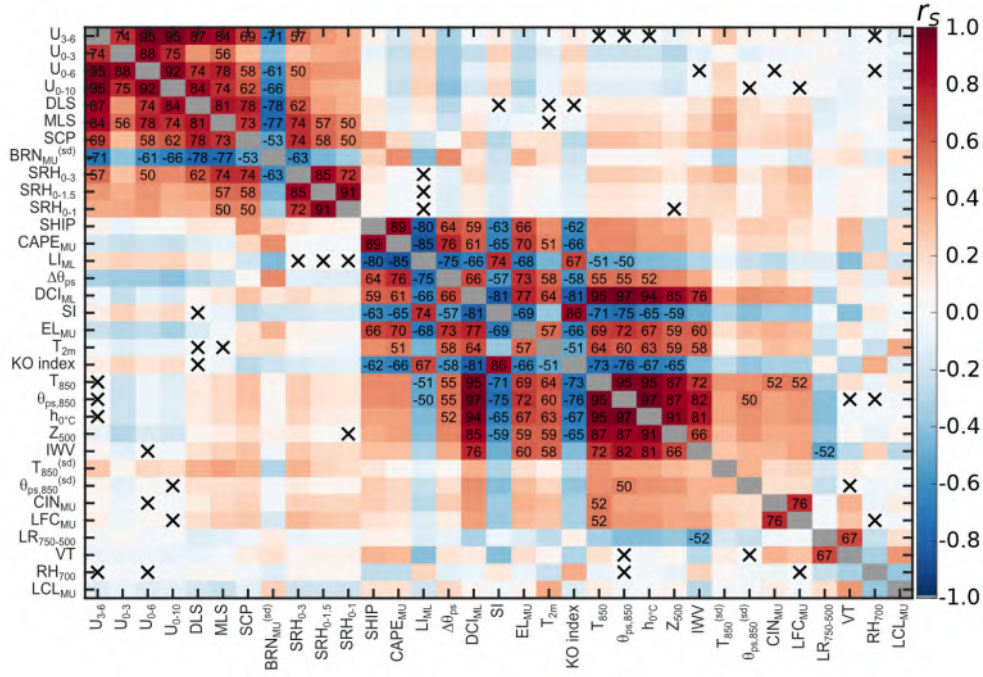
Figure 3.5: Correlation matrix of Spearman's rank correlation coefficient r_s. The definitions of abbreviations used can be found in the Table E.1 of Wilhelm (2022). Red color indicates high correlations ($r_s > 0$), while blue color indicates anti-correlations ($r_s < 0$). In addition, correlations with values above 0.5 are given as percentages (%). Statistically non-significant correlations are marked with a cross (Figure originally from Wilhelm, 2022).

and a temporal life-cycle average of the different ambient parameters are obtained for studying the correlation among each other by calculating the correlation coefficient matrix. Figure 3.5 shows Spearman's rank correlation matrix for the 33 ambient convective parameters used in the aforementioned study.

Note that the deep layer shear (DLS) is the primary determinant of the absolute value of the mid-tropospheric wind with stronger correlation coefficient ($r_s = 0.87$) with the mid-tropospheric mean wind between 3 and 6 km height above ground level U_{3-6}, whereas the correlation coefficient between DLS and the lower-tropospheric mean wind between 0 and 3 km U_{0-3} is much smaller. It has also been found that some of the supercell-related convective parameters have stronger relations to dynamical parameters. For example, the supercell composite parameter (SCP; Thompson et al., 2003), based on the multiplied product of CAPE, DLS and storm relative helicity (SRH), correlates more strongly with dynamical than thermodynamic parameters. The significant hail parameter (SHIP) composed of CAPE, DLS, water mixing ratio and middle layer lapse rate and temperatures, as another example, shows higher correlation with thermodynamic variables.

Areas of agglomeration in Figure 3.5 can be analyzed further using a non-hierarchical correlation clustering with the method of k-medoids-clustering shown in Figure 3.6, which is similar to the widely-used k-mean clustering. The distance metric is defined as $d = 1 - |r_S|$ (r_S, Spearman correlation coefficient) and indicates how much different variables correlate with each other. This clustering procedure enables a quick and concise overview of correlations and correlation-based distances between variables of interest. The red cluster represents mainly mid-tropospheric
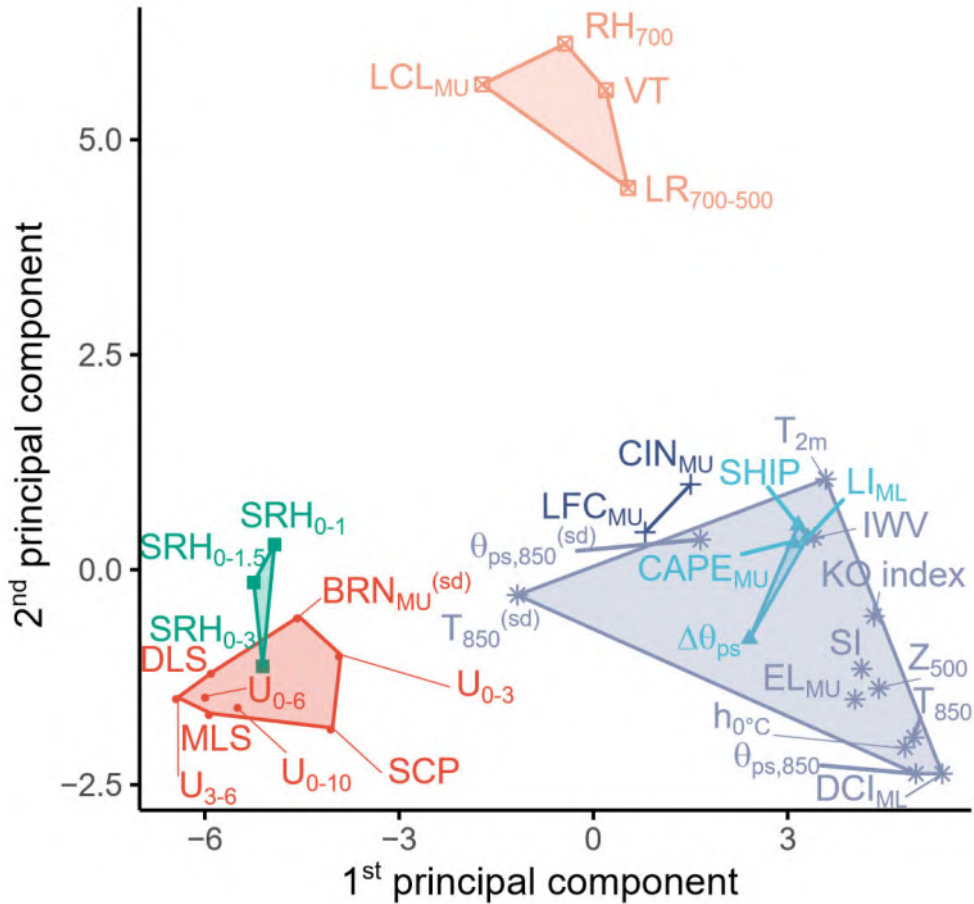
Figure 3.6: Based on a k-medoids clustering, $N_C = 6$ different clusters of ambient variables are represented. The dissimilation metric is defined as $d = 1 - |r_S|$. The clusters have been projected onto the first two principal axes of the eigenspace obtained from a multidimensional scaling analysis. The first principal axis accounts for 52.1 % of the observed variability, while the second axis accounts for 15.7% (Figure originally from Wilhelm, 2022).

dynamical quantities, whereas the dark blue cluster mainly shows thermodynamic parameters. The orange cluster contains mainly variables related to the mid-tropospheric instability. Note that this clustering also serves as a reference for selecting predictor candidates during the sensitivity tests of of the ML (see Section 3.1.4).

3.2.3 Mean absolute error

The root mean square error (RMSE) and the mean absolute error (MAE) are two different standard statistical metrics to measure, for example, model performance (Willmott and Matsuura, 2005; Chai and Draxler, 2014a). RMSE favors more in representing model performance than MAE when the error fitted in Gaussian distribution, while MAE is an unambiguous measure of an average error in a more natural way compared to RMSE (Chai and Draxler, 2014b). MAE is therefore used in this thesis to monitor and evaluate model performance due to the Gaussian q-distribution of hail events.

MAE is calculated by averaging the sum of the differences between the prediction values (x_i) and the observed values (x):

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |x_i - x| \qquad (3.5)$$

where n is the number of error and $|x_i - x|$ are the absolute errors at every grid point.

3.2.4 Data pre-processing

Data pre-processing is required for the later data mining and ML with the purpose to scale the data within the same range in the whole of data set, namely normalization (Ali et al., 2014). Normalization takes place by casting the data to the specific range, usually between 0 and 1, when there are big differences in the ranges of different data features. For output visualization, de-normalization is also needed after normalization applied, to present the data in its original form for interpretability. Moreover, a density-based weighting function is set up after normalization to allow manual selection of weights for input data in ML.

Z-score normalization

There are many ways to normalize the input data. In this thesis, Z-score normalization is used to scale data in the same range. Z-score normalization Z_i is defined as:

$$Z_i = \frac{x_i - \mu}{\sigma}, \qquad (3.6)$$

where μ and σ are the mean and standard deviation of the data set, respectively. Note that normalization does not change the structure and distribution of the data set, therefore, a q-distribution of hail events can not be changed to a normal distribution after data pre-processing.

Density-based weighting function

As a result of the q-distribution of hail events, the existence of a certain amount of the extremely large hail-affected areas can reduce model performance. To address this situation, where hail events do not ideally fit a normal distribution, adding a density-based weighting function (DWF) to the input data modifies the weights of the extreme values. DWF is aimed to improve the training of machine learning models for regression tasks that require more accurate prediction of rare target values compared to more common ones (Steininger et al., 2021). This is particularly relevant in cases where rare samples are of special interest, such as when estimating extreme precipitation events or hail events with a very large affected area. The parameter α is used to adjust the weighting scheme, with larger values of alpha emphasizing rare samples more strongly (while $\alpha = 0.0$ indicates uniform weighting across all samples). DWF determines the rarity of a target value by evaluating its density using Kernel Density Estimation (KDE).

3.2.5 Convolutional neural network

Some previous applications of Machine Learning for analyzing or prediction severe thunderstorms have shown promising results with higher prediction accuracy in the spatial and temporal scale of thunderstorms (Czernecki et al., 2019; Gagne et al., 2017, 2015). Among different models of ML, taking one of significant advantage of convolutional neural network (CNN) in studying convective storms is to rapidly capture and identify the certain features from the spatio-temporal meteorological data sets (Ren et al., 2021; Gagne II et al., 2019).

The advantage of using a CNN in this thesis is its outcome-orientated application, which offers an opportunity to directly investigate the potential relationship between hailstorms and ambient conditions including spatial information, rather than firstly to parameterize the complicated physical process between environmental conditions of hailstorm and characteristics of hailstorms in Germany. Here, different sets of candidate predictors in Table 3.2) with 3D ambient fields (longitude, latitude and date time, respectively) from ERA5 are determined. The CNN is also applied for sensitivity tests for candidate predictors to search for the best proxies of environmental parameters according to the lowest validation accuracy and validation loss during the training process.

The CNN in this study (see Figure 3.7) is set up with one consecutive convolutional layer, connected to a max-pooling and then a dropout layer followed after. This set of three layers is repeated once and another consecutive convolutional layer is added behind. All consecutive convolutional layers in this neural network have a 3×3 kernel size and use the ReLU activation function. The ReLU activation function introduces the property of non-linearity to a deep learning model and solves the vanishing gradients issue. Also, it only interprets the positive part of its argument, as the hail-affected area in reality is always positive. The first and last two convolutional layers have 32 and 64 filters, respectively. The network also includes two max-pooling layers, which is to aggressively downsample feature maps, with a 2×2 pooling size and stride of 2. Two dropout layers are used with dropout rates of 0.2. The dropout rate is the fraction of the features that are zeroed out. The first dense layer with ReLU activation and the second dense layer without activation function have 64 and 1 nodes, each separately. Also note that the optimizer 'Adam' is used during the training of CNN. According to Kingma and Ba (2014), Adam optimization is a stochastic gradient descent method, based on adaptive estimation of first-order and second-order moments. It is favorable for high computational cost and noisy sparse gradients within the data processing.

3.2.6 Stepwise feature selection and sensitivity test

Stepwise feature selection is a commonly used technique in machine learning and data mining. It aims to select a subset of relevant features that can increase the accuracy, performance, and interpretability of the model by reducing the dimensionality of the input data set. The process consists of three main branches: forward selection, backward deletion, and two-way deletion, in which elements are iteratively added or removed based on their contribution to the prediction action. This technique can be particularly useful in multivariate data sets, where the number of features
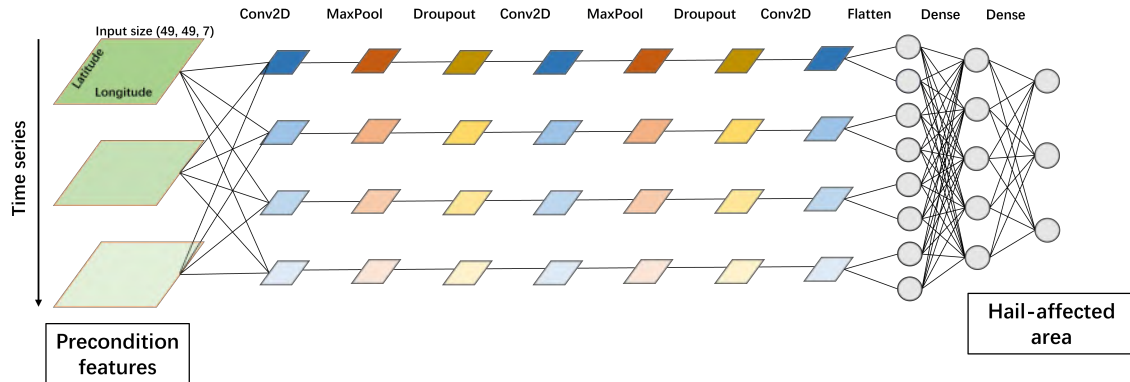
Figure 3.7: Schematic structure of convolutional neural network with different layers used in this Master's thesis and the number of nodes in this ML model is reduced.

is much larger than the number of cases. By selecting information features, incremental feature selection can improve model performance and interpretation by identifying the most important variables affecting forecasting activities. However, it is important to carefully evaluate the feature selection results and to ensure that the selected features are robust and representative of the underlying data distribution.

In this thesis, the forward selection method (Mohr et al., 2015) is conducted to find in the first step the best single predictor with the lowest MAE of model performance among all of the environmental parameters listed in Table 3.2. In the next step, the second predictor is determined by the lowest MAEs of the model output, adding one more parameter among the candidates and using the MAEs to monitor model performance. Further predictors are determined by continuously following the rule step by step (shown in Figure 3.8). If the accuracy of the model cannot be improved when adding one more parameter, feature selection ends. This process, so-called sensitivity test, also provides the information on model sensitivity as to which convective parameters are more beneficial to the forecast accuracy of the hail prediction and therefore should be considered as a potential proxy of hailstorms.

In an attempt to avoid confusion, the difference between "candidate parameters" and "predictors" is clarified in this paragraph. The candidate of convective parameters is highly related to convective storms and essentially play an important role for the development of hailstorms, but is not used in the final machine learning model. Predictors are chosen from the best result of the sensitivity test which confirmed the improvement of ML model prediction when using the stepwise feature selection method.

3.2.7 Class activation map

Although ML models are commonly viewed as "black boxes", there are techniques to visualize the learned representations in CNNs. One such technique is the gradient-weighted class activation
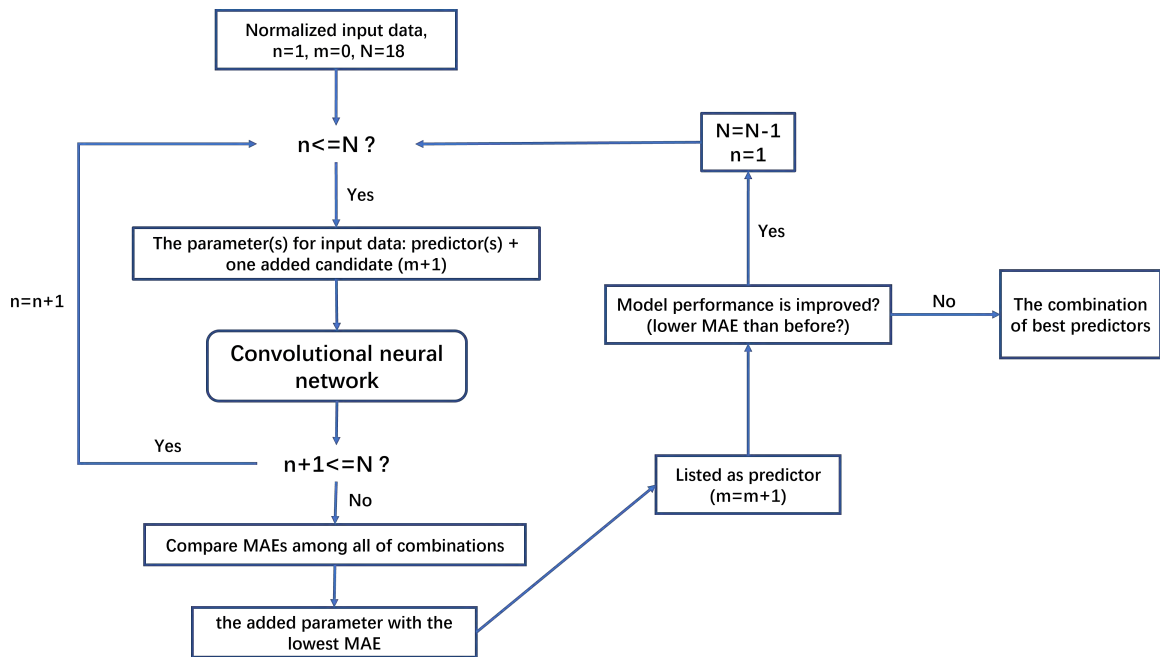
Figure 3.8: The flow chart illustrates each step of the stepwise feature selection and sensitivity test in details. $N = 18$ is the total number of candidates of environmental convective parameters, m is the total number of the chosen predictors which starts with zero.

mapping (Grad-CAM), which can help determine which parts of an input contribute to its final classification decision (Chollet, 2021).

Instead of using traditional RGB images, this thesis employs seven features that correspond to different colors, resulting in an image with seven float values at each timestep rather than three. Grad-CAM utilizes gradient information from the last convolutional layer to generate a coarse localisation map, which highlights important regions in the 2D input data for predicting how large the hail-affected area will be on the certain days. By overlaying the coarse localisation map with the features (Selvaraju et al., 2017), areas of particular importance can be visually identified. The resulting heatmaps show which predictors have more significant weighting in the combination of best predictors for model output, implying that the location of the variables are seemingly more important indicators for predicting hail storms to the machine learning model.

3.2.8 Reference prediction models

Two classic forecast models without any machine learning techniques are implemented as reference models to compare the prediction performance with machine learning models. One reference model is based on the climatological distribution of potential hail events, and the other model is to apply the persistence forecast on the daily potential hail events. Here, they are named as climatology forecast model and persistence forecast model.

Climatological reference forecast

The purpose of the climatological reference forecast is to predict long-term climate conditions, based on the analysis of historical data and statistics so as to determine the probability of future weather conditions over a given period of time, typically months to years beforehand (Murphy, 1992). It assumes that past climate conditions can be used to predict future weather under the assumption that the long-term climate remains stationary. Despite its potential benefits, the drawback of this model in reducing the large uncertainty of model performance is still the main challenge, as the model only relies on many simplified assumptions of the atmosphere, and is therefore heavily influenced by errors and biases. Here, the climatological reference forecast is the daily mean of the potentially hail-affected area averaged over the 13-year period (2005–2017), to predict daily hail-affected areas during 2018–2019. It is smoothed by a 31-day moving window. The skill of the reference forecast is compared to the ML model forecast by calculating the MAE between forecast and observations.

Persistence forecast model

Short-term weather patterns can be predicted using persistence forecasting by assuming that weather conditions persist for a few days. The principle of persistence forecasting is the continuity of atmospheric motion and weather patterns with little or no change in the overall weather pattern, and therefore, the current conditions are likely to continue for a period of several hours to days. The challenge for persistence forecast is mainly that it does not account for the sudden change in weather and the influence of large-scale atmosphere patterns, which can significantly affect local weather conditions. Here, 1-day ahead hail-affected area is used to predict the hail-affected area on the current day for the test periods. Therefore, the MAE between 1-day ahead forecast and observation in the same time periods (2018-2019) as climatology model is computed as well to estimate the performance of model.

3.2.9 Categorical verification

It has been common practice to verify non-probabilistic forecasts for discrete variables (Murphy, 1996). Categorical verification involves predicting a categorical or discrete variable in order to assess the accuracy of a forecast. To determine the quality of a forecast, accuracy, bias, and various skill scores are calculated by comparing forecasts to observations. It is widely used in fields such as prediction of precipitation, cloud cover, or severe weather events. The data sets are entered into a 2×2 contingency table after determining an appropriate threshold. This threshold here refers to a specific value of the potentially hail-affected area (shown in Section 5.3). The elements $a - d$ are classified in Figure 3.9 based on whether an event was observed (Yes/No) and whether it was predicted (Yes/No). It shows the relationship between the counts of forecast/event pairs for dichotomous non-probabilistic verification (Wilks, 2011). However, it can also be used for the evaluation of thunderstorm quantities, as applied, for example, by Doswell et al. (1990) or Kunz (2007).

In categorical validation, a hail event is deemed to have taken place or not, based on a predefined threshold. The crucial aspect of this validation process lies in identifying the optimal threshold that

		Prediction	
		Yes	No
Observation	Yes	**a** correct forecast	**c** surprise events
	No	**b** false alarms	**d** non-events

Figure 3.9: The elements in the 2×2 contingency table provide an understanding of how forecasts and observations are related in non-probabilistic verification situations.

maximizes the occurrences of correct forecast a while simultaneously minimizing the number of occurrences of false alarms b and surprise events c.

Many different scalar attributes have been created and utilized to describe the performance of forecasts based on the contingency table (Figure 3.9). The widely used scalar attributes to comprehensively evaluate the forecast model performance are listed below.

Accuracy

The accuracy values indicate the correspondence between the forecast and the event intended to predict. A perfect forecasting in the contingency table displays the situation $b = c = 0$, meaning that Yes/No forecasts perfectly match the Yes/No observations in the reality (Wilks, 2011). Therefore, as the most intuitive measure to estimate the proportion correct (PC), namely accuracy, is defined as:

$$PC = \frac{a+d}{n},\tag{3.7}$$

where n is the sample size ($n = a+b+c+d$). It is seemingly not an ideal attribute especially when Yes events are infrequent in the sample (Wilks, 2011).

Critical success index (CSI)

The Critical success index (CSI) as an alternative to the PC is particularly applied when the correct forecast for the Yes observations appears less commonly than non-events (Wilks, 2011). It is also sometimes called *Threat score* (TS) and defined as:

$$CSI = TS = \frac{a}{a+b+c}.\tag{3.8}$$

In many studies, the CSI is utilized to evaluate spatial forecasts that are issued simultaneously, such as warnings for severe weather (Doswell et al., 1990; Ebert and McBride, 2000; Kunz, 2007).

However, it does not provide an unbiased indication of forecast skill, but rather shows a proportional relationship with the frequency of the event to be forecasted (Schaefer, 1990). Therefore, more forecast skill scores need to be taken into account while using CSI at the same time.

Probability of detection (POD)

Detection probability (also known as hit rate, See Eq. 3.9) is a measure of how often accurate correct forecasts are made relative to the total number of the Yes observations. Alternatively, it can be described as the proportion of situations happened in which the predicted event actually occurred and was also for the forecast.

$$POD = H = \frac{a}{a+c} \tag{3.9}$$

Heidke skill score (HSS)

The Heidke skill score (HSS) measures the proportion of correct forecasts obtained by random forecasts which are statistically independent of the observations. It is widely applied in the evaluation of thunderstorm-related phenomena (Kunz, 2007; Gagne II et al., 2019; Czernecki et al., 2019). A perfect HSS value receives a score of 1, a total random forecast gets a score of 0, and a worst forecast gets a score of -1 (Heidke, 1926; Wilks, 2011).

$$HSS = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)} \tag{3.10}$$

Peirce skill score (PSS)

The Peirce skill score (PSS) is a measure that involves subtracting two conditional probabilities in the likelihood-based rate factorization of the joint distribution between POD and the probability of false detection (POFD). It is akin to the HSS but differs in that the denominator features the reference POD for unbiased, random forecasts (Wilks, 2011).

$$PSS = \frac{ad-bc}{(a+c)(b+d)} \tag{3.11}$$

False alarm ratio (FAR)

It is also crucial for the forecast to quantify the uncertainty and bias of model prediction. False alarm ratio (FAR) is the fraction of the miss forecast that should predict yes but fail in the situation

with the Yes observations. FAR is commonly used for analyzing the forecast performance skill and defined as:

$$FAR = \frac{b}{a+b}.$$

(3.12)

Probability of false detection (POFD)

In an attempt to analyze the discrimination skill of the model forecast, the probability of false detection (POFD), also named as False alarm rate (F), discriminates incorrect forecast for the non-events:

$$POFD = F = \frac{b}{b+d}.$$

(3.13)

Noted that the definitions of FAR and POFD are sometimes confusing, therefore it is necessary to clarify the difference. The FAR quantifies how many cases the Yes observations in reality are not successfully predicted by the model forecast, evaluating the ability of the model to predict correct forecast, while the POFD assess how bad the ability of model is to identify correctly the non-events in the observations.

4 Statistics between hail events and ambient condition

The development of a hail prediction model based on ML, a statistical analysis of 15 years of radar-based hail track identification and the corresponding environmental conditions with 18 candidates of convective parameters in Germany is a necessary first step to investigate their possible relationships. Section 4.1 gives an overview over the 15 years of hail frequency, the trend of hail occurrence, the hail seasonal cycle and also their characteristics during this period. In Section 4.2, the mean fields of the hail-related convective parameters are further analyzed to identify spatial features favourable for hailstorms. In the next Section 4.3, correlation maps between the hailstorms and convective parameters in different regions of Germany are calculated in order to select appropriate candidates for the next step of setting up the machine learning model.

Note that "hailstorms" in this thesis are not directly the hail-produced thunderstorms recorded by radar, but the identified hail tracks by the TRACE3D algorithm (Section 3.1.2) potentially develop to the thunderstorms that can produce the hail (Section 3.1.3).

4.1 Overviews of hail tracks

Hail frequency

Hail frequency as one of the most important features of hail-producing thunderstorms is firstly analyzed to study their annual variability (Punge and Kunz, 2016). The most direct approach involves quantifying the number of instances of hail that take place annually in a particular location (Eulerian perspective). In this thesis, the hail frequency in Germany is statistically represented by the number of hailstorms (hail tracks) per year.

Figure 4.1 illustrates the total number of hailstorms recorded per year using the radar-derived hail data set from 2005 to 2019 from a Lagrangian perspective (see Section 3.1.2). The slightly increasing trend in the 15-year hail climatology in Germany (red line in Figure 4.1) is very weak and not statistically significant, although the annual variability is relatively large among each year. Note the pronounced drop in 2012 with the lowest number of hail events caused by a technical problem from the radar stations during that year, so that several potential hailstorms were not recorded.
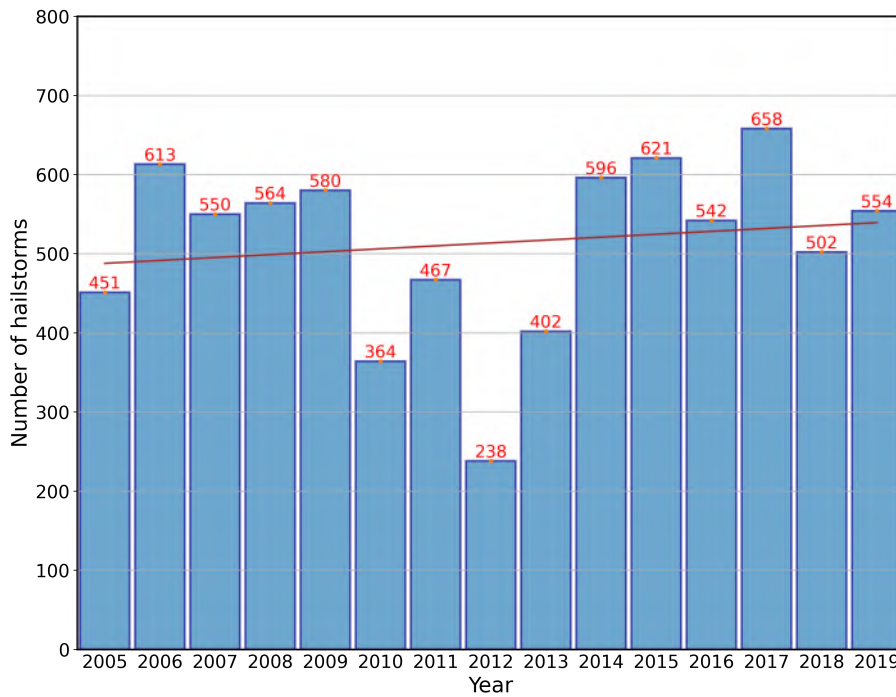
Seasonal cycle

Figure 4.1: Annual frequency of radar-derived pontential hailstorms for the period 2005-2019. The red solid line is the linear regression for the number of hailstorms per year.

Hail season is defined by the historical statistics rather than a clear definition, mainly due to the variety of hail characteristics (e.g., hail-affected area, maximum size of hailstones, hail accumulation) and regional discrepancies. Based on earlier studies, for example, by Gudd (2004); Punge and Kunz (2016); Kunz et al. (2018, 2020), hailstorms typically occur between April and September in Germany. Figure 4.2 shows the cumulative daily average of radar-derived potential hail events that occurred during the hail seasons with a 14-day moving average window. Two significant peaks of hail events appeared in June and July, which is in agreement to the radar-based analyses of hailstorms by other authors (Gudd, 2004; Kunz and Puskeiler, 2010). There is also a relatively high probability for hailstorm occurrence in early August and September under suitable conditions. Besides, since 2015, hailstorms occurred more frequently in ealier June. The reasonable explanation could be an earlier-starting summer season due to the temperature increases regarding climate change. It is subject to further investigations, but not the purpose of this study.

Characteristics of hail tracks

After determining the certain periods by data availability for the ML model set-up to predict the hail-affected area, the target to represent the hailstorm-affected area is determined by analyzing the characteristics of hail tracks derived from radar data. During the study period from 2005 to 2019 a total of 7702 hailstorms on 911 days were identified. On 1591 days no hail days (i.e., no track identified in the radar data) were registered. Figure 4.3 shows the distribution of different hailstorm-related parameters. An approximate Gaussian distribution for the propagation speed of the hailstorms overlaps the result with a peak at 26–38 km h^{-1} (about 25% occurrence) and a minimum at ≥ 98 km h^{-1} (of $> 1\%$ occurrence in total). Meanwhile, roughly 67 % of the hailstorms lasted 45–90 minutes. The width of a hailstorm is determined by identifying the largest
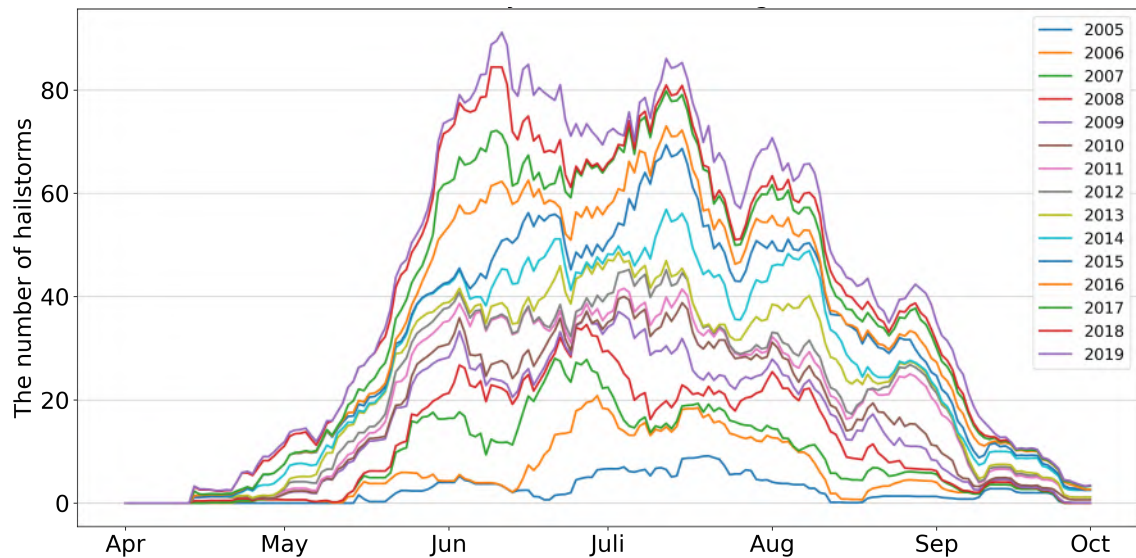
Figure 4.2: Cumulative daily hail occurrence with an 14-day moving average of radar-derived hail days during the hail season in Germany.

reflectivity core ($\geq$ 55 dBZ). Over 61 % of all potential hailstorms have a width of 3–8 km. The results may be over- or underestimated after averaging, but may still reflect the characteristics of the hailstorms to some extent. Estimates of the area affected are based on a function of the length and width of the hail tracks (Section 3.1.3) to quantify the spatial area affected by potential hailstorms. More than 52% of the hail-affected areas for the single haistorms is of up to 400 km^2, while it is not in the agreement with the mean area of 20.5 km^2 from the ealier research of Changnon (1977). More severe events may result in larger affected areas in the approximate rectangle shape of the tracks.

In general, any features of hail tracks can be considered as a prediction target for the ML model. Here, the hail-affected area rather than just the width or length of hail tracks or others are chosen to represent the potential hailstorm threat for two reasons. Firstly, assuming a positive relationship between the potential hailstorm threat and hail track length or width, it may be exceptional, for example, that some severe hailstorms are shorter in length but wider in width. Secondly, while the estimated area of these rectangles may introduce some level of uncertainty, their larger values could still potentially show a correlation with the higher potential hailstorm threat. Moreover, these rectangles can provide the local quantitative information, serving as a practical reference for insurance companies. At the same time, before building up the machine learning model, it still remains an open task to explore the convective environmental conditions that may be conducive to the development of hailstorms.

4.2 Comparison among environmental mean fields

Since hail-producing convective storms are relatively rare even during the hail season, the spatial-averaged mean fields on convective pre-conditions based on the ERA5 reanalysis data are obtained
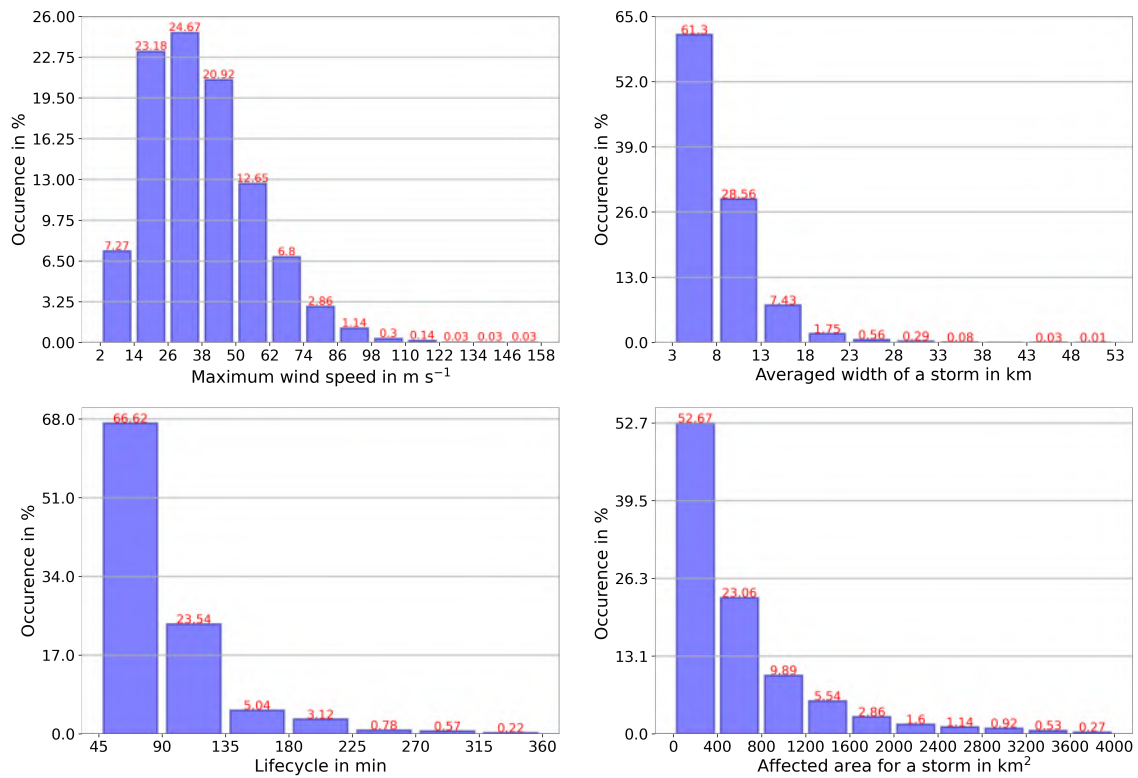
Figure 4.3: Histograms of different parameters of hail tracks from 2005 to 2019. The left top (bottom) is the storm propagation speed (lifetime) during tracking of the single hailstorms; The right top is the averaged width for single hailstorms and the right bottom is the estimated affected area for single tracks based on Section 3.1.3.

beforehand for 18 different potential candidates in order to better zoom in on the general convective conditions related to hail occurrence.

Here, CAPESHEAR is exemplarily discussed in detail (Figure 4.4) as it turns out to be the best single predictor in the ML model sensitivity tests (Section 6.1). In addition, this quantity also estimates the likelihood of well-organized and long-lived convective storms (Section 3.1.5). The diurnal cycle of convection is primarily influenced by near-surface temperature fluctuations and the subsequent changes in the atmospheric boundary layer (Punge and Kunz, 2016). Key triggers, such as low-level convergence due to thermal circulations in complex terrains, land use variations, or coastal land-sea breezes, also intensify in relation to the diurnal temperature cycle, although occasionally with a time lag (Markowski and Richardson, 2011). According to Bedka (2011), hailstorms in Europe are more likely to occur in the afternoon and early evening (See Figure A.1, Schmidberger, 2018), with some variations depending on terrain features and proximity to the coast. To match the remote sensing measurement of hailstorm detection, 12 UTC (14 CEST, central European summer time) is the empirical time point for representing the pre-convective conditions of hailstorms in the radar-based hail data set (Kunz and Puskeiler, 2010; Mohr and Kunz, 2013).

Figure 4.4a illustrates the mean field of CAPESHEAR during the hail season (April - October) from 2005 to 2019. The area with larger CAPESHEAR is mostly located in southern Germany, especially in Baden-Württemberg and the southern part of Bavaria, while pre-convective conditions are weaker in the northern part of Germany. It is because the lower temperature and less moisture

leads to the lower CAPE associated with CAPESHEAR in the northern Germany (Section 3.1.5). Another possible explanation for this is that the general climatology of weather conditions for the hail season diluted the pre-conditions of hailstorms in the northern Germany, although hailstorms are more frequent during these periods (Section 4.1).

Therefore, the mean field only for hail days (identified tracks) is shown in Figure 4.4b to focus only on the convective conditions prior to hail occurrence. Higher CAPESHEAR values can be seen in northern, central and southern Germany, but the spatial trend from north-to-south remains similar, which is also consistent with the statement that the hail frequency in Germany increases from north to south (Punge and Kunz, 2016). Also note that when hail events occurred, there would be the higher CAPESHEAR in the environmental conditions during the hail season (See Figure 4.4c).

To further test our assumption that larger affected areas are associated with higher potential hailstorm threat, the hail-only situation (Figure 4.4b) is therefore subsampled and divided into 3 terciles (large, medium, small) based on the values of the affected area by daily accumulations. Figure 4.4d shows the lower tercile with small daily affected areas, while Figure 4.4e is upper tercile with large daily affected areas. One interesting finding is that CAPESHEAR with small daily affected areas is much weaker than that with large affected areas, which is in agreement with our expectations. Another point is that orographic features like mountains and slope terrains in Baden-Württemberg, Bavaria, and also northern Hesse (Figure 4.4e) exhibit seemingly a higher convective predisposition with higher CAPESHEAR values, presumably because of the higher wind shear in complex terrain. The prominent features with high CAPESHEAR overlap the potential hail hot spots in Figure 4.4f. Flow deflections in horizontal and vertical directions caused by mountains can trigger thunderstorm initiation, including for those producing hail, and help them to persist in favorable conditions. Therefore, topographical features of a region can greatly influence the frequency of hailstorms, with mountain ranges, especially locations downstream, usually being the most common location for hail hot spots(Punge and Kunz, 2016; Fluck et al., 2021).

However, the individual contribution of either CAPE or BULKSH to the CAPESHEAR is more dominated by the distribution of CAPE (Figure A.2, Figure A.5, Section 3.1.5). Some mean fields of parameters such as CAPE, SHIP, TT (Figure A.2, A.4, A.3) show a north-to-south gradient similar to that of CAPESHEAR. The candidates that offer the general atmospheric conditions during the hail season as background, such as U500, T850, BULKSH (Figure A.5) do not show such a north-to-south gradient difference between the mean field distributions.

The investigation of the climatological environmental conditions of hailstorms partly show the potential connection that more favorable convective conditions can lead to the higher potential hailstorm threat. However, the analysis only showed the general pre-conditions of hailstorms and lacks the sufficient quantitative evidence related to the hail-affected area.
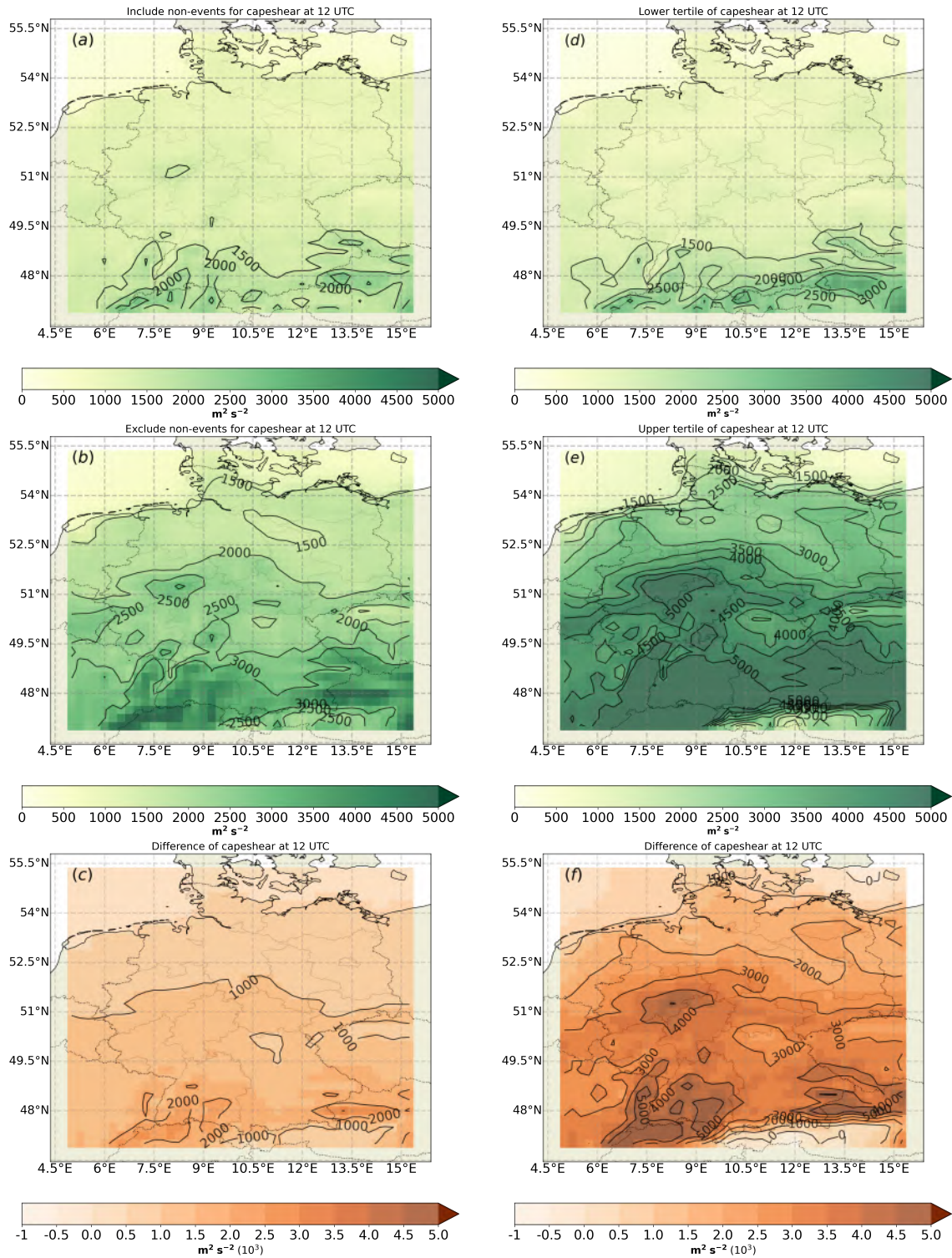
Figure 4.4: Mean fields of CAPESHEAR during the hail season from 2005-2019. (a) is the mean field including the dates with non-hail events happened, while (b) is the same as (a) but excluding the dates without hail occurrence; (c) is the result of (b) − (a). For the further sub-sampling under the condition of (b), (d) is the mean of just the dates below the lower tercile ranking of the affected area by daily accumulation, and (e) is just in place of the dates above the upper tercile ranking like (d); (f) is the difference of (d) − (e). The contoured black lines in (a,b,d,e) only highlight where the values of CAPESHEAR larger than 1500, indicating the tendency of very active convection (Grieser, 2012).

4.3 Correlation maps between convective parameters and hailstorms

To further investigate the link between hail tracks and convective parameters for the optimization of the ML model, all 18 atmospheric parameters listed in Section 3.1.5 are correlated with the daily hail-affected area (Section 3.2.1). Noted that the correlation analysis here does not correspond to the location of the individual hailstorms, but rather to the temporal correlation analysis between convective parameters and daily hail-affected area in Germany at a certain location. Therefore, if there are high correlation coefficients at certain locations, the correlation analysis is relative reliable, but conversely, no or weak correlations do not seem to be a complete indication that the parameters are not related to hailstorms.

CAPESHEAR, as the most interesting candidate, is exemplarily discussed in detail, rather than showing the spatial features of correlation for all the convective candidates. Here, the daily affected area from the centre location of the hailstorms (Figure 3.4) is taken separately for different regions (Northern: $\geq 52.0°N$; Southern: $\leq 49.5°N$; Central: in between) of Germany, in order to avoid diluting the spatial information of hail hot spots when considering Germany as a whole. In addition, the non-hail days are also added in the samples because the ML model should be able to identify the pre-conditions of hailstorms from the general weather conditions individually.

In northern Germany (Figure 4.5a), several areas with higher correlation coefficients are located in the north-east of Lower Saxony and also in the east of Northrhine-Westphalia. This indicates a moderate relation between CAPESHEAR and the daily hail-affected areas. Meanwhile, the region on the eastern edge of Northrhine-Westfalia, with high hail-affected areas in the central Germany (Figure 4.5b) overlapped somewhat with the areas in the northern Germany, but with a slightly stronger relationship. In Figure 4.5c, the higher correlation values for the region are located in Baden-Württemberg and Bavaria in the southern region of Germany, where can potentially observe the higher CAPESHEAR (Figure 4.4), and the higher daily hail-affected area .

In fact, several studies have found indications that suggest that areas with enhanced hail hazard show a connection to specific terrain features (Giaiotti et al., 2003; Kunz and Puskeiler, 2010; Punge and Kunz, 2016). Note that the areas of the high CAPESHEAR convective environments that occur in the mean field (Figure 4.4b) correspond to some extent to the geographical distribution of hot spots with the higher daily hail-affected area in the correlation maps in the three different sub-regions of Germany (Figure 4.5).

However, the correlation between hailstorm-favoring environments and hailstorms is somewhat weakened, when CAPE (Figure A.6) or BULKSH (Figure A.7) is considered separately. It seems that the q-distribution of CAPE carry with BULKSH mainly dominates the correlation with hail. Other storm-orientated convective candidates such as SHIP (Figure A.8), with more or less the same range of maximum correlation coefficients, show a similar positive correlation distribution to CAPESHEAR. Likewise SLI shows a relative strong negative correlation with the daily affected area associated with hailstorms (Galway, 1956), based on the fact that low SLI values represent

high instability of the atmosphere (Section 3.1.5). Other candidates for the general atmospheric conditions such as wind speed U500, temperature T850, humidity Q850 remained poorly associated with hail occurrence. The possible reason could be the q-distribution of the daily area affected to represent the hail occurrence, leading to a higher correlation with the similar distributions of convective parameters, such as CAPE, SHIP, etc.

For the further preparation of the machine learning setup, the correlation maps as a reference can give a hint which single convective parameter potentially contribute the most to the machine learning model. Note that the maximum values of the correlation maps within the moderate limits are mainly influenced by the fact that most of the non-hail days (defined in Section 4.1) dominate the general weather conditions. Moreover, parameters such as BULKSH, U500 and VV6km, which do not have a significant correlation with the hail affected-areas according to the present analysis, still may play an important role locally for the occurrence and development of hailstorms and, thus, shall also be considered in the ML model.

In addition to the correlation analysis between ambient conditions and hail-affected area, our selection of candidates also considers the independence among convective parameters from different clusters, which represent dynamic, thermodynamic and mid-tropospheric instability quantities, respectively, according to Wilhelm (2022) in Section 3.2.2. Choosing independent parameters enables the machine learning model to capture the variety of environments of hailstorms in different aspects as much as possible.

(a) Northern Germany

(b) Central Germany

(c) Southern Germany
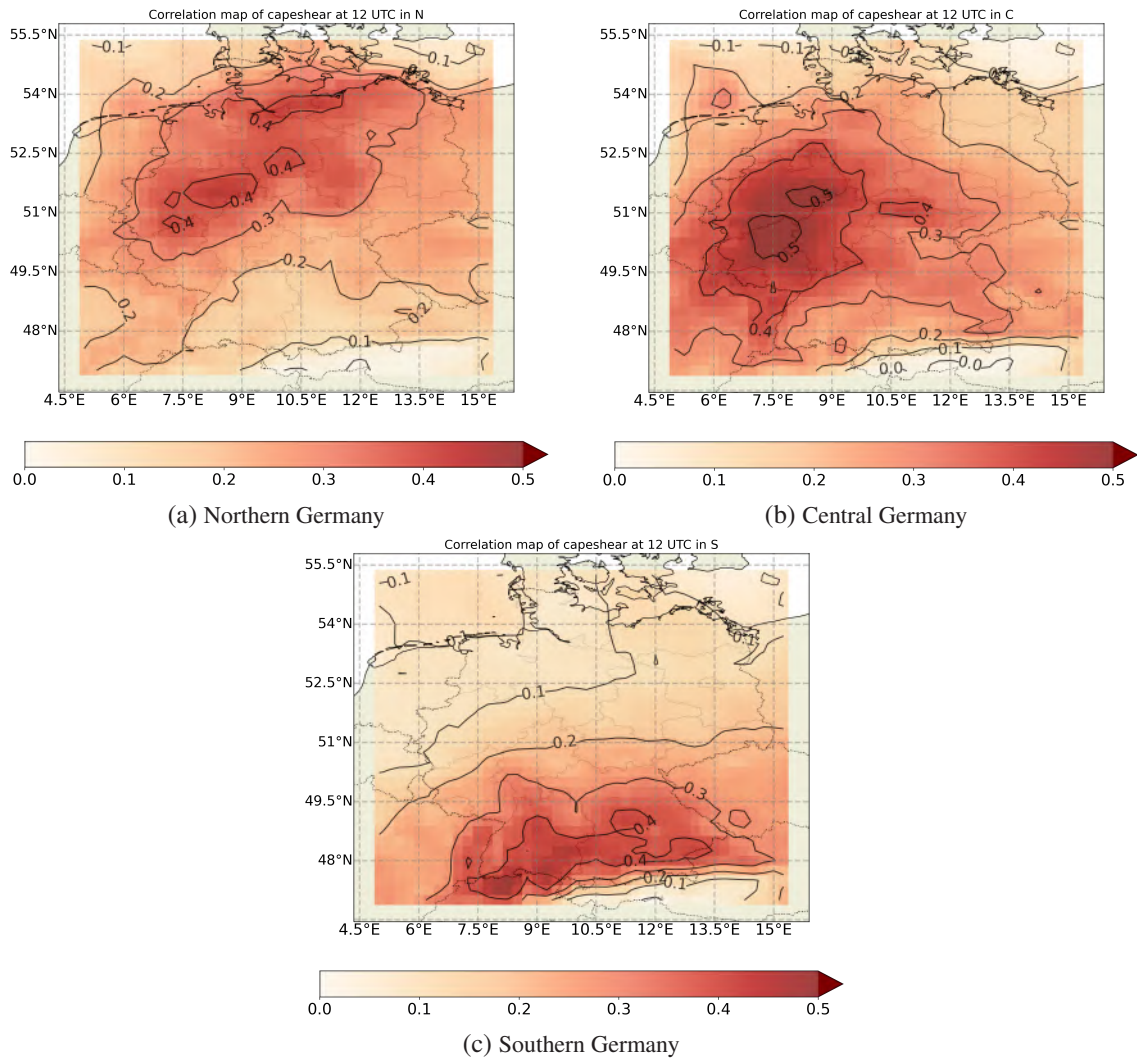
Figure 4.5: Correlation maps between CAPESHEAR and the daily affected area at 12 UTC during the hail season from 2005 to 2019 including non-hail events.

5 Evaluation of ML model performance on hail prediction

The integration of disparate data sources from reanalysis data and the radar-based remote sensing data, in conjunction with the implementation of machine learning techniques, represents a promising approach for augmenting the dependability of large hail-affected area, with the high possibility to surpass the performance of any individual product that is typically employed in the context of operational forecasting at the present (Czernecki et al., 2019). Here, the convolutional neural network (CNN) is implemented to predict the daily affected area for hailstorms. The training is conducted only during the hail season (April to September) from 2005 to 2017 and 2 years were used as validation data from the 13 years of training data. The 2 years of test data are also taken during the hail season, but from 2018 to 2019. The performance of a machine learning model is always monitored using the Mean Absolute Error (MAE) to quantitatively measure the difference between model prediction and observation.

In the following Section 5.1, the result from ML output is firstly presented and the adjustments of the model set-up are discussed as well. Section 5.2 presents the performances of the non-machine learning models in the same test period as a reference, and also other adjustments of the ML model set-up to compare the model prediction ability of the machine learning model. The evaluation of the machine learning output for the deterministic forecasts of rare binary events using the categorical verification is explicitly discussed in Section 5.3. Finally the potential limitations to achieving the goal of this study using the machine learning are well justified in Section 5.4.

5.1 Output and adjustment of the ML model

As mentioned in the earlier Sections 3.2.4 and 4.3, the q-distribution, or the uneven distribution of hail events across different magnitudes, poses a challenge to ML models in accurately forecasting the occurrence of extreme events. To address this issue, an essential consideration lies in the allocation of greater significance to infrequent hail events in the input data processing. To this end, an auxiliary weighting function known as the density-based weighting function α (Section 3.2.4) is incorporated into the CNN.

Figure 5.1a shows the daily affected area predicted from the convolutional neural network in the test period, where $\alpha = 0.0$ means that a uniform weighting across all samples during the training periods is decided. The graph highlights that ML models show relatively improved MAEs in predicting non-hail events, while struggling to achieve comparable performance in predicting extreme hail

events, particularly those of exceptional spatial extent (daily affected area $\geq 20,000\,\mathrm{km}^2$). These challenges might stem from a lack of adequate data and the non-linear relationship between hailstorms and environmental conditions.

From Figure 5.1b, assigning a value of 0.5 for α does not significantly improve the ability of the ML model to accurately predict very severe hail days in terms of large affected areas that exceed the $20,000\,\mathrm{km}^2$ threshold. Nevertheless, the predictions for the large spatial extend of the affected areas ($\leq 20,000\,\mathrm{km}^2$) show a moderate level of improvement and a greater closeness to the actual observations. By gradually increasing the assigned value up to α=1.5 (Figure 5.1c), the machine learning model achieves improved detection of the affected area of hail events, but at the cost of significantly overestimating the prediction of non-hail events. This overprediction of non-hail events has a largely negative impact on the overall MAE of the prediction.

Thus, the weighting function for the machine learning model should have a dual capability, encompassing an increased predictive performance in identifying non-hail events, while at the same time achieving a minmum degree of MAE in predicting the affected area associated with hail events. Given the aforementioned observations in Figure 5.1d, the α value of 0.5 was considered optimal, as it yielded better model performance with the lower MAEs in subsequent training rounds. At the same time, the MAEs remained stable for the original structure of input data for the ML model. The values ≥ 1.0 for α already degrade the performance of the ML model and therefore should not be used.

While Figure 5.1 exhibits an encouraging visualization of the outputs from the machine learning model, the determination of its superiority and its ability to compare favorably with alternate forecast models remains inconclusive. It therefore requires extensive further research.

5.2 Comparison of different model results

To compare the performance of ML models with traditional non-machine learning forecasting models, and the climatology forecast and the persistence forecast as reference models, respectively, Table 5.1 provides explicit evidence that the predictive ability of the ML model is superior to the other two models. The lowest averaged MAE of ML reflects its higher predictive ability. Notably, the improved performance is associated with remarkably low MAEs, further demonstrating its higher predictive skill among these three models.

Name of model	Range of MAEs	Averaged MAE
CNN	$0.20 - 0.30$	0.22
Climatology model	$0.40 - 0.82$	0.63
Persistence model	$0.46 - 0.50$	0.48

Table 5.1: Summary of the performance measures for all models (CNN, climatology and persistence model).

Figure 5.1: The outputs of the convolutional neural network. (a)-(c) affected area predicted for environmental conditions at 12 UTC in the test period (2018-2019) during the hail season with different values of the density-based weighting function α; the blue solid line is the ML prediction and the black dashed line is the observation for comparison. (d) is the validation MAEs of model performance as a function of training epochs during the model training with different values of α.

On the other hand, it is crucial to delve into the intricacies of refining the settings of the machine learning model in order to assess its optimal performance capabilities. The granular details of these settings can potentially hold the key to extracting the maximum predictive power from the model.

12 **vs.** 18 UTC:

The 12 UTC is empirically determined to represent the prevailing hailstorm environment (Mohr and Kunz, 2013), while 18 UTC typically corresponds to when the nighttime hailstorms begin to form. The model performance between these two times is compared to test the better ability of the model to predict the hail-affected areas at both times. This is due to the potential for significant changes in hailstorm environmental conditions within this time period (Punge and Kunz, 2016), which could act as an important signal to the machine learning model.

Figure 5.2a provides a comparative performance attribute of the machine learning models designated as 12 and 18 UTC during their respective training phase. It is noteworthy to observe that the 12 UTC model exhibits a marginally superior performance in predicting the area of extreme large hail-affected area relative to its 18 UTC counterpart (Figure 5.2b). However, the disparity in the data distribution between these two model predictions with respect to the observed values do not show significant deviations (Figure 5.2c).

Based on these findings, the rationale for using the 12 UTC machine learning model for hail prediction in this study has been substantiated. This choice is largely influenced by the observation that the majority of storms identified as hail by radar detection systems are predominantly manifested during daylight hours $(12 - 18$ UTC), seen in Figure A.1 in the Appendix. Consequently, the environmental preconditions represented by the 12 UTC model are closer in time to the occurrence of the hail event. This proximity ensures that the prediction is reasonably close to the actual observed value.

Note that radar records essentially capture the threshold of the radar reflectivity larger than 55 dBZ within storm events. For those thunderstorms recorded after 12 UTC, the predominant trends and signals within the environmental conditions are often possibly discernible earlier before 12 UTC.

Non-linear activation vs linear activation:

Note that in the training process of the above machine learning model, nonlinear activation functions (ReLU) are employed to diminish the interdependence between input data from different layers of the model and prevent excessive cluttering of data information from interfering with the learning capabilities of model (Pedamonti, 2018; Section 3.2.5). However, the study of Wilhelm (2022) summarized the presence of correlations among various types of environmental convective variables, including dynamical, thermodynamical, and atmospheric variables. Therefore, this thesis also undertakes an investigation into the impact of these correlations on the predictive sensitivity of machine learning models. This investigation involves the use of linear activation functions within machine learning models at 12 UTC. it is found that the use of a linear activation function provided machine learning models with limited information about hailstorm environmental conditions to predict hail-affected areas, while introducing more instability and model bias with high MAE (see the suddenly negative values in Figure A.10).

Thus, the inclusion of non-linear activation functions becomes indispensable as it enables the network to take into account non-linear associations between hailstorms and environmental conditions during the decision-making process. While there is a possibility that correlations among environmental parameters could potentially provide valuable information regarding hailstorms for machine learning, the introduction of the non-linear activation function remains imperative to adequately capture and model the complex relationships involved.

5.3 Categorical verification of ML model predictions

Categorical verification offers a systematic approach for evaluating the predictive accuracy of ML-based hail prediction models and determining suitable thresholds (Wilks, 2011). The datasets are organized into a 2×2 contingency table (Figure 3.9 in Section 3.2.9). In the earlier study of Kunz (2007), various convective indices are quantitatively compared to investigate the preconvective environment and to assess the predictive skill of the convective parameters and indices in forecasting thunderstorms and related impacts such as hail or flood damage.
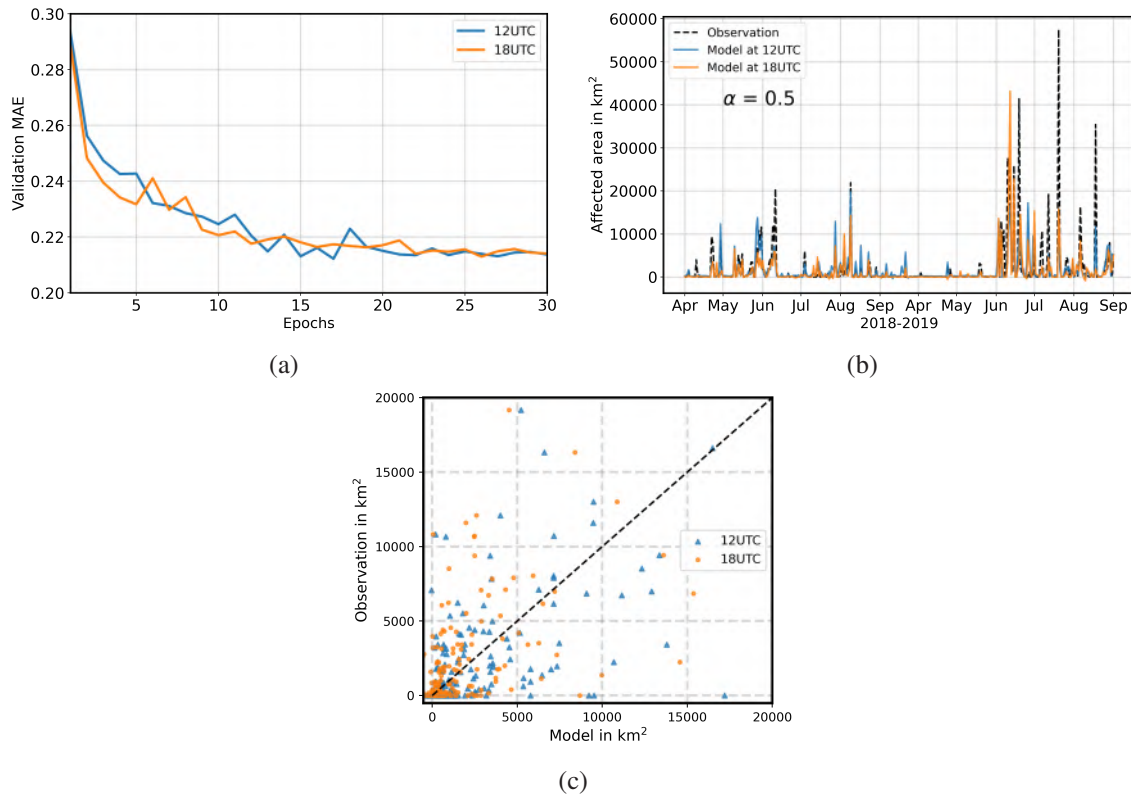
(a)

(b)

(c)

Figure 5.2: The results of the ML model for 12 (the blues) and 18 UTC (the oranges). (a) is shows the validation MAEs during the training; (b) shows the daily affected area predicted in the test periods; (c) is a scatter plot between observations and predictions.

In contrast, this study focuses on directly evaluating the accuracy of ML-based hail predictions in relation to the actual hail target generated by the model. As the magnitude of the potential hailstorm threat can directly relate to the strength of the hail-favorable environment on a daily basis (Section 4.2). It is important to note that the daily affected area of hailstorms refers to the region potentially at risk of hail, rather than the actual ground area experiencing hail (Section 3.1.3). Thus, it becomes necessary to establish a specific threshold to accurately identify the occurrence of hail events in radar-based scenarios. The selection of this threshold directly impacts the predictive capability of the machine learning model for hail prediction.

Furthermore, the process of identifying appropriate thresholds requires an initial quantification of the uncertainty within the machine learning model. By eliminating any inherent biases in the model, the resulting threshold outcomes are more likely to align with the actual value of the hail-affected area. The different skill scores depicted in Figure 5.3 demonstrate the variations observed in performance after reducing model biases. The threshold for optimal performance was determined based on a comprehensive analysis of the 1740 km^2 daily affected area by hailstorms. The evaluation criteria included maximising the Heidke skill score (HSS) and Peirce skill score (PSS) while minimising the false alarm rate (FAR). At this threshold, HSS was 0.66, PSS was 0.69 and FAR was only 0.3. For the same threshold of the hail-affected area, HSSs of climatology forecast and persistence forecast were 0.51 and 0.36, respectively. The accuracy of the model was even found to be 88% with POD = 0.78 and CSI = 0.58. The slightly superior mode performance at 18 UTC is listed in the Appendix for the purpose of comparison (Figure A.11, Figure A.12).
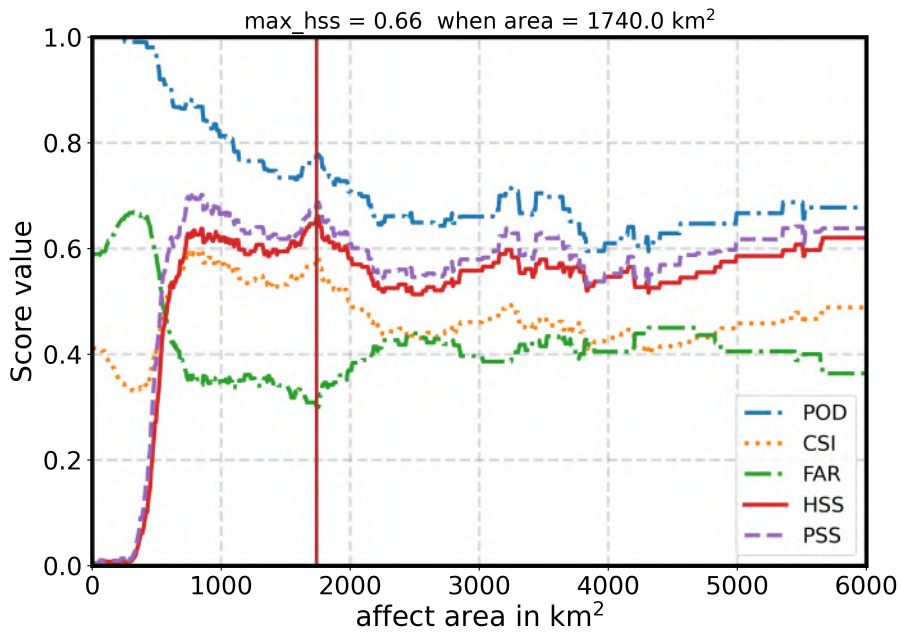
Figure 5.3: Different skill scores as a function of the daily affected area by hailstorms.
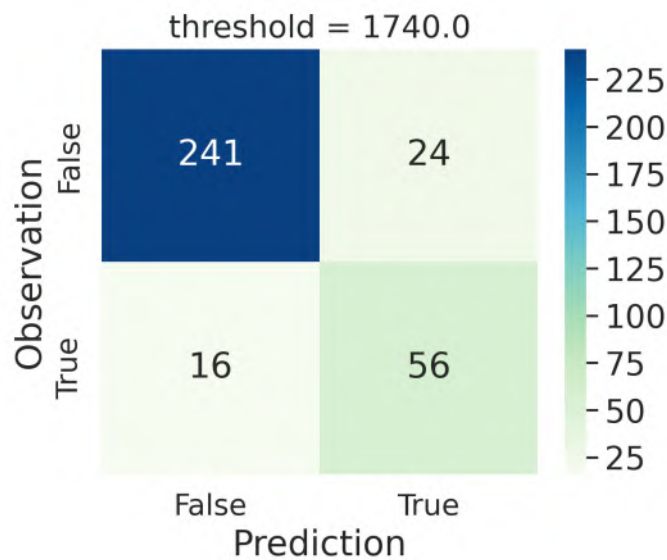
Figure 5.4: Contingency table for the optimal threshold of 1740 km^2 for the daily hail-affected area.

However, as already discussed in Section 5.2, the application of the model still focuses on 12 UTC because of the potential for the prediction purpose based on the pre-condtion.

In addition, detailed information about the number of observed and predicted events for each of the four cases of the contingency table (a-d) with the optimal threshold can be found in Figure 5.4. It is attempted to shed light on the core of the machine learning model by investigating how it succeeds or fails in predicting hail events, with a particular focus on gaining knowledge of the environmental preconditions associated with hail occurrence (correct forecast a and surprise event c) in Section 6.2 of the next Chapter.

5.4 Potential limitations

Despite achieving commendable performance in prediction skill scores, machine learning is not yet a perfect tool for hail forecasting. The accuracy of machine learning predictions is highly dependent on the availability and quality of convective environmental field information included in the input data. Hence, it is essential to further investigate the potential limitations that may arise regarding the predictive ability of the machine learning model, particularly in relation to the structural aspects of the input data.

First, the predictive performance of machine learning is directly influenced by convective environmental variables, which are carefully selected based on statistical correlations and the previous research (Wilhelm, 2022). While statistical methods have demonstrated correlations between atmospheric ambient conditions and hail, certain variables such as BULKSH and VV6km exhibit relatively low correlations, indicating unsatisfactory connections with hail (see Section 4.3). This is primarily attributed to the q-distribution of hail events, underscoring the need to consider factors beyond correlation coefficient analysis when identifying suitable parameters. Additionally, limitations in data availability prevent the inclusion of highly correlated environmental variables like SRH or KO-index (Kunz, 2007; Ukkonen et al., 2017), potentially hindering the optimization of machine learning model performance. Inserting these variables may hold the promise for further enhancing the predictive capabilities of the machine learning model.

Second, the current model exclusively examines the convective field at a specific time point, leading to a temporal bias in a portion of the hailstorm environmental field predictions. Consequently, this bias can have a significant impact on the accuracy of the machine learning predictions. For example, consider a scenario where a SCS with a lifetime of only one hour develops at 11 UTC. In such a case, the environmental field information available at 12 UTC may not provide the machine learning model with the relevant data needed to make accurate forecasts. By incorporating the environmental field corresponding to the active time period of the daily hailstorms into the machine learning model, the dynamic changes in the evolution of convective storms are captured within the the input data. This implementation can be promisingly expected to also improve the prediction performance of the machine learning model.

Moreover, the daily-accumulated hail-affected areas do not include specific geographical information regarding the occurrence of hail. This limitation arises due to the hourly time steps of the reanalysis data, making it challenging to accurately quantify the location of individual hailstorms while matching them to the corresponding reanalysis data in the early stages of machine learning model development. However, as the model progresses, there are opportunities for optimisation. Once a sufficiently large volume of input data is available, machine learning models can effectively analyze and comprehend the relationship between other distinguishing characteristics of hailstorms and the associated preconvective environmental fields.

Above all, while there remains scope for future optimization and improvement of machine learning model, the current findings indicate a relatively superior model performance and favorable prediction outcomes using the CNN.

6 Interpretation of convolutional neural networks for hail predictors

The selection of predictors tries to cover convective parameters included in the clusters of Figure 3.6 (Section 3.2.2) and also those with high correlations with hail-affected areas (Section 3.2.1 and 4.3). On the other hand, it is often perceived that deep learning models are "black boxes" as learning representations are difficult to extract and present in a human-readable form (Chollet, 2021). A deep understanding on how CNNs make self-decisions to predict hail-affected areas, based on information gathering and processing of different predictors, is still untouched and the focus of this chapter.

The sensitivity test results in Section 6.1 reveal the predictors that are optimal for ML-based predictions of the area affected by potential hail. Section 6.2 introduces a visualization approach with case studies that highlights the distinctive characteristics of 2D preconvective fields that drive machine learning models to identify hailstorms in specific environmental contexts, especially the correct forecasts and false alarms. Section 6.3 attempts to extract potential feature patterns of these individual hailstorms from the environmental fields. These feature patterns may possibly be the indicative signal to amplify and capture environmental information about future hailstorm development, further improving hail forecast accuracy in turn.

6.1 The combination of best predictors

To discern the optimal environmental parameters that yield hail prediction outcomes superior to traditional approaches, while simultaneously maintaining relatively low computational costs, the stepwise feature selection method (Details in the Section 3.2.6) is applied to a pre-selected set of 18 candidate convective parameters. As a result of this selection process, seven predictors that significantly improve the machine learning performance are successfully determined. The details of these identified predictors and their contribution towards optimizing the prediction results is illustrated in Figure 6.1.

The machine learning model identified CAPESHEAR as the most important single predictor for hail, outperforming other convective parameters such VT and TT, primarily due to its responsiveness to both the instability indices VT, TT and wind shear, as depicted in Figure 6.1a. The synthesis of humidity Q850, mid-latitude zonal wind speed U500, BULKSH, and instability indices (VT, TT) related to convective storms, collectively contributed to the best performance of the model. This performance was observed to increase progressively with the number of predictors, particularly

after the inclusion of the second best predictor, Q850, in which moisture is covered, but not for CAPESHEAR. This addition yields a relative improvement of 6% in the model performance, as illustrated in Figure 6.1b.

Note that most of the important predictors are dynamic quaties rather than thermodynamic ones, although VV6km, U500 and BULKSH obviously only slightly improve the MAE of the model performance. In the study of Wilhelm (2022), there is a larger correlation between BULKSH and U500 (Figure 3.5), but BULKSH has the very low correlation with hail-affected area (Figure A.7). Thus, less additional environmental information on development of hailstorms that can be provided for the machine learning model. However, this still reflects the effect of wind shear and the impact of the large-scale circulation in the middle troposphere on small to meso-scale convective systems that generate hail (Mohr et al., 2020).

The combination of the best predictors at 18 UTC is obtained from the sensitivity test in parallel, as it can also be used for the prediction of nighttime hailstorms. The results are accompanied by comparable values of MAEs and a relatively stable performance that represents a model performance essentially competitive with that for 12 UTC, except that the order, in which the predictors were selected, was different. Also the moisture component is replaced by relative humidity RH850. It is also found that CAPESHEAR for 18UTC alone achieves relatively similar MAEs for the combination of seven predictors (see Figure A.13 in Appendix).

Overall, the application of a stepwise feature selection methodology has facilitated the efficient identification of the most advantageous predictors by machine learning algorithms. These predictors improve forecast accuracy while taking into account three crucial aspects (Figure 3.6) that govern convective storm development: dynamical parameters, thermodynamic parameters, and thermodynamic-related instability indices which is highly associated with thermodynamics. However, it remains an open question which spatial information contain most information for the machine learning models. Thus, further analyses and exploration are imperative to deepen our understanding in this direction.

6.2 Analysis of activation heatmaps

Because of the randomness in the data during the training process, the order of the released best predictors does not necessarily represent the importance of the machine learning decisions on a daily prediction. Heatmaps of class activation is regarded as a useful tool to identify which input play a significant role in the decision-making process (Section 3.2.7). It is a suitable diagnostic to specifically visualize where more favourable environmental conditions prevail for hailstorms for the ML model to predict the daily hail-affected area. In addition, the visualization outcome is very helpful for analyzing the causes of false alarms in machine learning prediction for further model optimization (Chollet, 2021).

Out of a total of 72 observed hail events shown in Figure 5.4, one instance is selected from each of the following two categories for further study: 56 events where the machine learning model
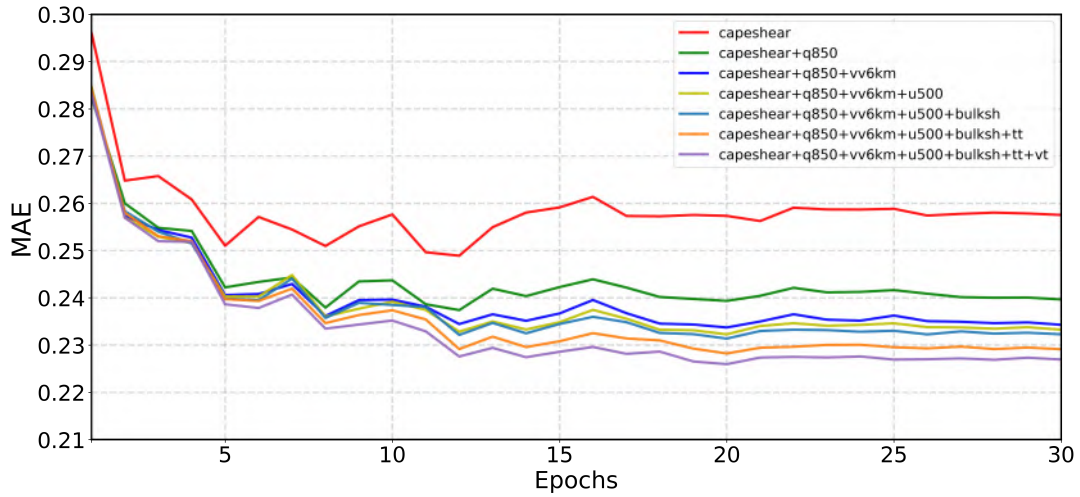
(a) The variation of MAEs as a function of the number of iterations during training for the sensitivity test.
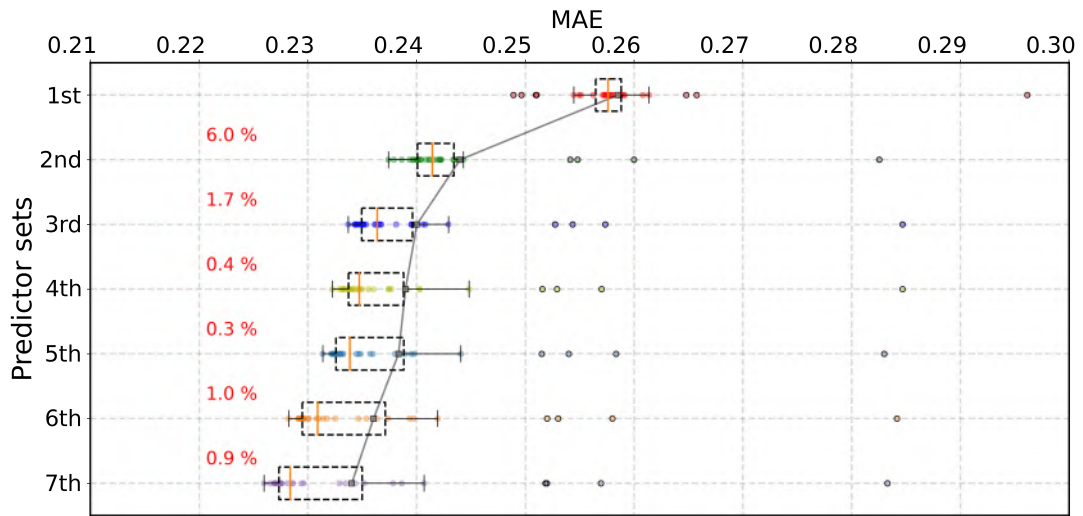
(b) Boxplot of the MAEs distribution in each training period. The Grey line is the variation of averaged MAE of each training. Percentage numbers represent the relative increased rate of MAEs compared with the last training.

Figure 6.1: Model performance for preconditions at 12 UTC, represented as the measurement of MAEs from standardized training data during the sensitivity test using the method of the stepwise feature selection. Predictor sets of Y-axis in (b) correspond to the legend labels of (a).

accurately predicted hail (hit cases), and 16 events where the model's prediction was incorrect (miss cases) based on the categorical verification (Section 3.2.9). This selection is made with the intention of visualizing the distinct contributions of various convective predictors to the machine learning model and to delve into an analysis of the reasons behind the failures in machine learning predictions.

Figure 6.2a shows the ambient conditions of the 7 convective predictors for the hail event successfully predicted by machine learning on 13th, May, 2018, as well as the specific location of the area affected by hailstorms on that day (pink polygons in the lower right subplot of Fig. 6.2a). At that time, a weak easterly flow related to a blocking anticyclone on its south-eastern flank prevailed over Germany, causing strong wind shear in the vertical configuration of the atmosphere and resulting in a cluster of hailstorms (Mohr et al., 2020).

It is found that the red contours of the activation heatmap, when superimposed on top of the ambient field, mirrored and covered the spatial allocation of CAPESHEAR more directly and inclusively. Especially the region with the highest activation matches the locations where hailstorms actually occurred. This observation also holds true for other successfully predicted hailstorms, suggesting a direct relationship between the spatial distribution of CAPESHEAR and the predictions of the machine learning model. This also substantiates the identification of CAPESHEAR as the most effective single predictor in the stepwise feature selection process.

The distribution of other convective predictors such as VT, TT, and other environmental fields contribute differently to the activation heatmaps for each hail event, but the area with larger values of different predictors always overlappe the distribution of the relative strong activation, depending on the intensity of each predictor on the day and the specific environmental factors.

Note that the machine learning model receives instantaneous fields of predictors, while its targets are daily totals of hail-affected area, which did not include any information about the exact location of the hailstorms. However, the ability of machine learning to target the location of hailstorm clusters based on the different importance of the environmental fields of the predictors is quite promising, as seen in the activation heatmaps.

Under the same condition of large-scale atmospheric circulation, the machine learning model failed to accurately predict the hail events that occurred on 15th, May, 2018 (Figure 6.2b). On that day, the hailstorms manifested themselves mostly in the north-eastern region of Germany. Despite the presence of high values for BULKSH and VV6km in close proximity to the hailstorms, the environmental conditions were less than ideal. This is primarily attributable to the low values CAPESHEAR in the vicinity of the hailstorms. As a result, the machine learning model was misguided by the low values CAPESHEAR and unable to make an accurate prediction. The highest activation occurs in the southwestern corner where CAPESHEAR is relative high, leading to the occurrence of the incorrect prediction, despite the prevalence of beneficial moisture conditions (Q850) and atmospheric instability (VT, TT), which were predominantly concentrated in central Germany.

Comparing the manifestation of hailstorms in the two German case studies discussed above, it can be seen that the hailstorm events on 13 May occurred predominantly in relatively restricted geographical zones, mainly between 12 and 16 UTC. In contrast, on 15 May, the day characterised by an incorrect forecast, hailstorms, although also mainly occurring between 12 and 15 UTC, had a comparatively wide geographical distribution and were of short duration. This anomalous pattern, to a certain degree, elucidates the inaccuracies encountered by the machine learning model in its predictive ability on that particular day.

Nevertheless, the aforementioned findings derived from the two case studies ascertain the proposition that machine learning possesses the potential ability to accurately anticipate the emergence of large aggregations of SCS and may effectively contribute to daily hail prediction for short lead times. If combined with NWP output the CNN has a high potential to be applied for short-term to medium-range forecasts in the future.
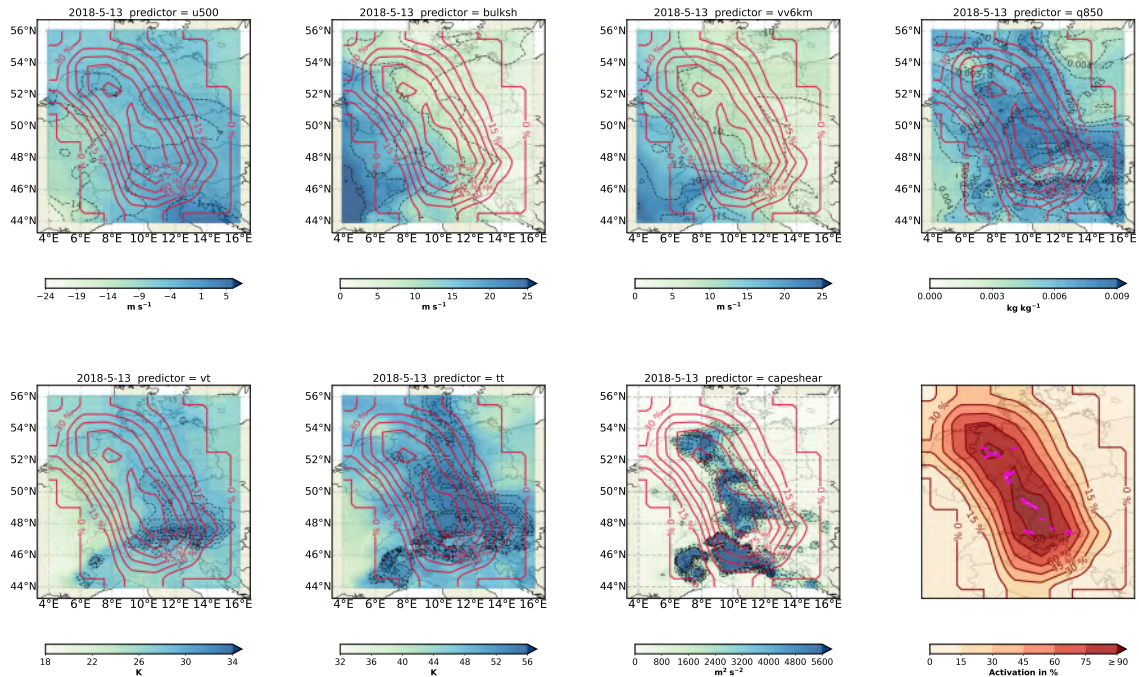
(a) Case study of correct forecast on 13th, May, 2018. Convective predictor is titled in each subplot, except the lower right subplot with pink polygons, which shows only the location of the affected areas by the corresponding hailstorms and the shaded heatmap of class activation relavant to CNN model.
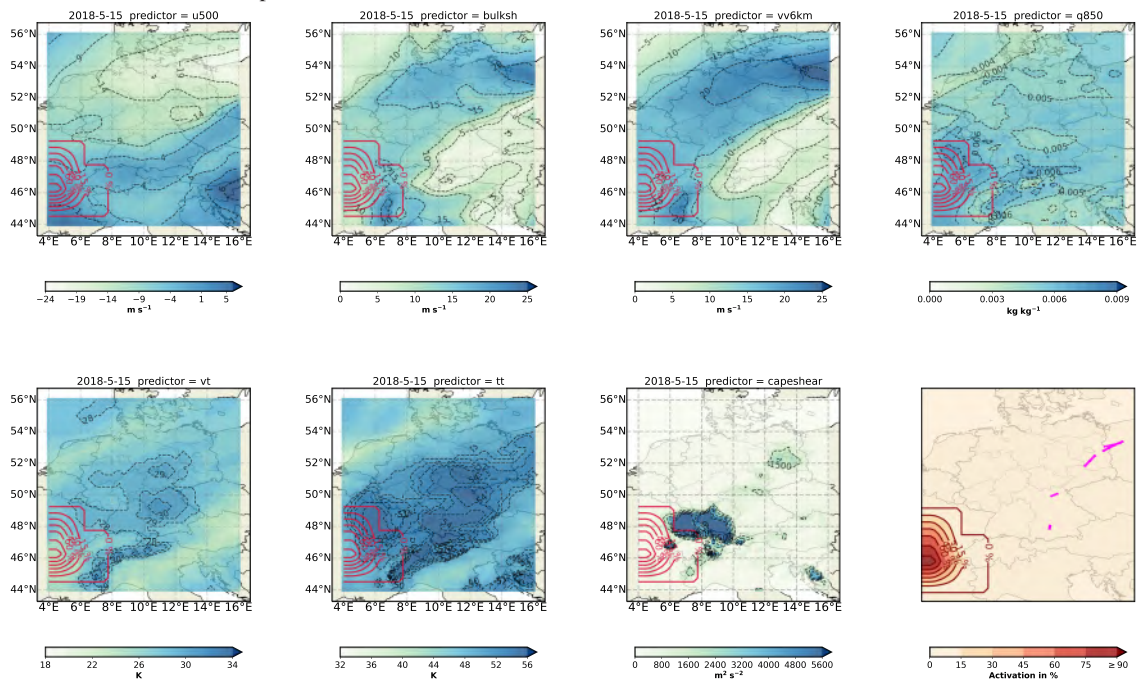
(b) Case study of False alarms on 15th, May, 2018. The detail shown in the subplots is the same as (a).

Figure 6.2: Different ambient conditions (blue shading) at 12 UTC for two case studies of (a) correct forecast and (b) false alarms from the outcomes of ML model. For VT, TT and CAPESHEAR only contours of the values are favorable for the active convective and the development of thunderstorms. The activation (in %) of the CNN (red shading and solid lines contours every 15%) demarcates regions that hold significant relevance for the CNN decision-making and is overlapped on the fields of convective predictors by red solid lines. Pink polygons in each lower right subplot represent the hail-affected area by every single hail track on that day.

6.3 Contribution of different predictors to heatmaps

Upon thorough comparison of cases from correct forecast and false alarms in machine learning-based prediction, it is found that when clusters of hailstorms manifest within the environmental domain, machine learning models tend to exhibit a higher likelihood of detection. Consequently, an intriguing query arises concerning the feasibility of applying a statistical approach, to identify the specific convective patterns prevalent in hailstorms (Kunz et al., 2020). This process is realized by averaging the corresponding activation heatmaps, which denote the decision importance assigned by the machine learning model. This exploration aims to foster a deeper comprehension of the interrelationship between hailstorm and their corresponding environmental conditions. Additionally, in the context of false alarm events, the utilization of the statistical analysis proves beneficial in elucidating the underlying factors responsible for inaccuracies in machine learning predictions. Such insights can subsequently facilitate the optimization of the model capabilities.

To achieve this goal, a total of 56 cases of accurately predicted hailstorms and 16 cases of false alarms, along with their corresponding activation heatmaps, were statistically analysed. These heatmaps represent regional averages over this period of time (Figure 6.3). The results indicate that there is no significant difference in lower-middle level wind and shear (BULKSH, VV6km) and convective indices between the two categories, as observed in the environmental fields.

However, it is noteworthy that the correct forecast cases (Figure 6.3a) exhibit a relatively high CAPESHEAR, together with a prominent large-scale westerly zonal wind at mid-level atmosphere over Germany (Mohr et al., 2019). Conversely, the false alarm cases (Figure 6.3b) lack these distinct features and display lower values of Q850, associated with the less humid atmospheric conditions.

Nevertheless, it is also found that after averaging activation heatmap still coincides to some extent with the distribution of the large value areas of the environmental field predictors. Although the activation heatmap of correct forecasts reveals a relative clear structure, the false alarm does not seem to have the clear and well-defined patterns for the prevailing predictors (the maximum of the averaged activation heatmap is only 35% for the correct forecast cases and 20% for the false alarm cases). One possible explanation for no significant maximum activation in hit cases is that the hail-affected areas in some false alarm cases were still moderately high but slightly below the optimal threshold, leading to the incorrect forecasts of ML model (Section 5.3). Another interesting point is regarding the weather situation between these two categories. For U500, which can reflect the large-scale atmospheric waves propagating in the upper atmosphere, there is a stronger westerly flow over Germany in the hit cases, associated with the northern flank of blocking systems or ridges. In the false alarm cases, the easterly flow prevail over Germany, associated with the northern flank of a low pressure systems.

This also mainly indicates a considerable degree of randomness in the geographical distribution of hailstorm occurrences, which leads to a significant regional weakening and dilution of the potential patterns from the averaged activation heatmap towards the conditions of hailstorms.
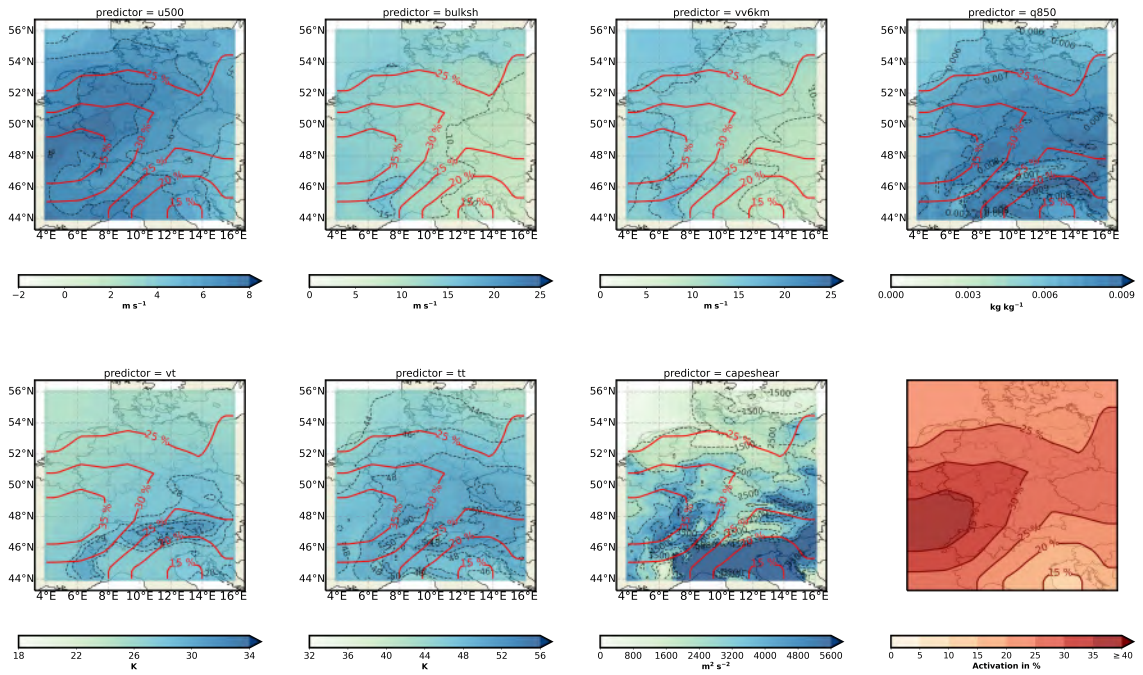
(a) Mean conditions of convective predictors for 56 correct forecast events.
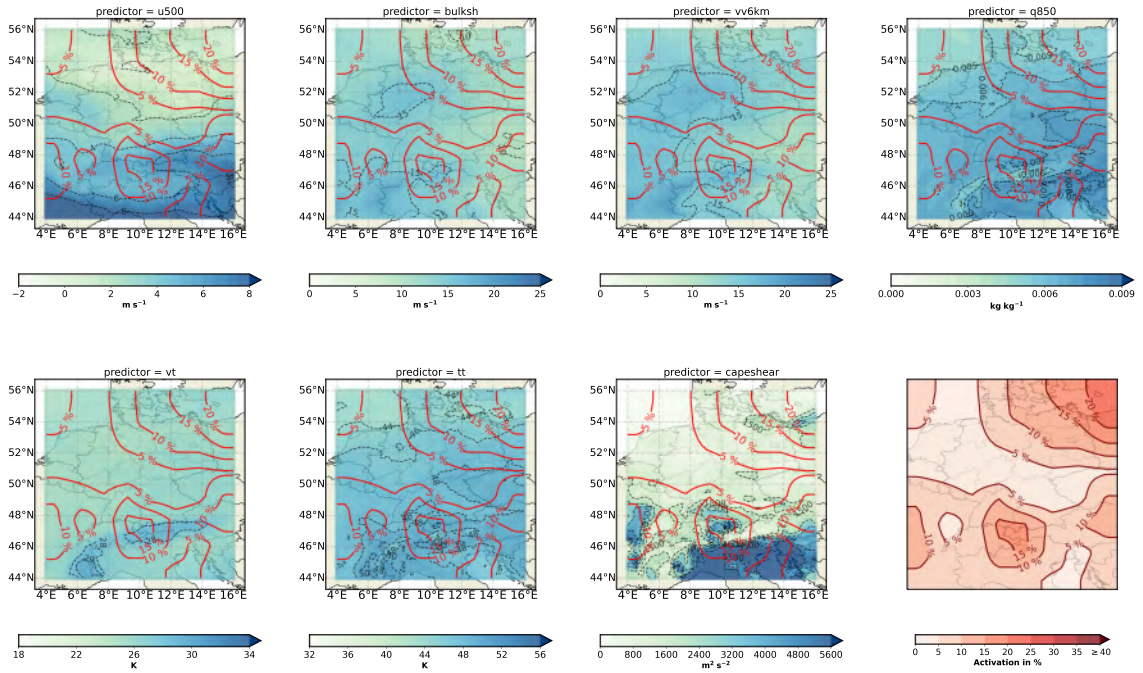
(b) Mean conditions of convective predictors for 16 false alarm events.

Figure 6.3: Mean fields of ambient conditions (blue shading) overlapped with the averaged activation heatmap (red solid line contours in every 5%) of (a) a total of 56 correct forecast events and (b) a total of 16 false alarm events in Figure 5.4. Details is the same as Figure 6.2.

The environmental mean field of hailstorms does not directly represent the connection between the activation heatmaps and the individual predictors. In order to further quantify and evaluate the importance of the daily predictors for machine learning predictions, the spatial correlations were calculated between the activation heatmaps and the environment fields of the predictors. The aim of this process is to quantitatively compare the contribution of the individual predictors of correct forecasts and false alarms for the machine learning decision-making (see Figure A.14 in Appendix for the specific correlation coefficients for every individual hail events).

For the two cases studied in Section 6.2, the correlation matrix shows that the activation heatmap of 13th, May, 2018 has relatively high correlation coefficients of 0.33, 0.57, 0.43, 0.61, and 0.46 with U500, Q850, VT, TT, and CAPESHEAR, respectively. The corresponding correlation coefficients for 15th, May, 2018 are relatively weak, with only 0.3, 0.39, -0.11, 0.12, and 0.14. This also reflects to some extent statistically how machine learning autonomously relies on environmental information to predict hail events.

Upon examination of the hail events categorized into correct and incorrect cases, it becomes evident that there exists considerable variability in the correlation coefficients between the different predictor fields and the heat maps (Figure A.14). Consequently, the outcomes are shown in a statistical boxplot (Figure 6.4]. The degree of aggregation of the correlation coefficients ranges in the figure also reflects to some extent the reliability and stability of machine learning for the individual predictors for hail prediction. For instance, CAPESHEAR, which exhibits a high correlation with activation heatmaps and a narrower distribution, reconfirms its potential as a reliable predictor for machine learning applications within the environmental conditions. In contrast, the correlation coefficient of U500, representing the importance of zonal winds for the machine learning prediction, showcases a broader distribution and thus increased variability. This is primarily because upper-level atmospheric circulation such as blocking can indirectly affect the formation of convective storms by influencing the atmospheric vertical configuration (Mohr et al., 2019, 2020). The complexity of these intermediate physical processes, which cannot be directly apprehended by machine learning, also contributes to the corresponding observed variability.

In addition, the generally higher heatmap correlation coefficients of VT, TT, and CAPESHEAR for the correct forecast events compared to the false alarm events are also a more intuitive reflection of the potential role and importance of the three parameters in the successful integration for the ML model predictive decisions. This further suggests that one direction in improving the accuracy of machine learning predictions may lie in how to amplify or increase the signal or importance of relatively important predictors when pre-processing the input data for ML model.

Nevertheless, the entire sample of 72 hail events in the test data is still insufficient, and the results obtained lack some degree of representativeness of the hailstorm environment due to the retention of some uncertainty, so more hail data and other methods are needed to confirm the conclusions. Moreover, activation heatmaps are useful to help diagnose model prediction errors. A more fine-grained and targeted analysis of heatmaps to identify problems requires a more appropriate approach for in-depth research. For example, heatmaps centered on individual hailstorms rather than the
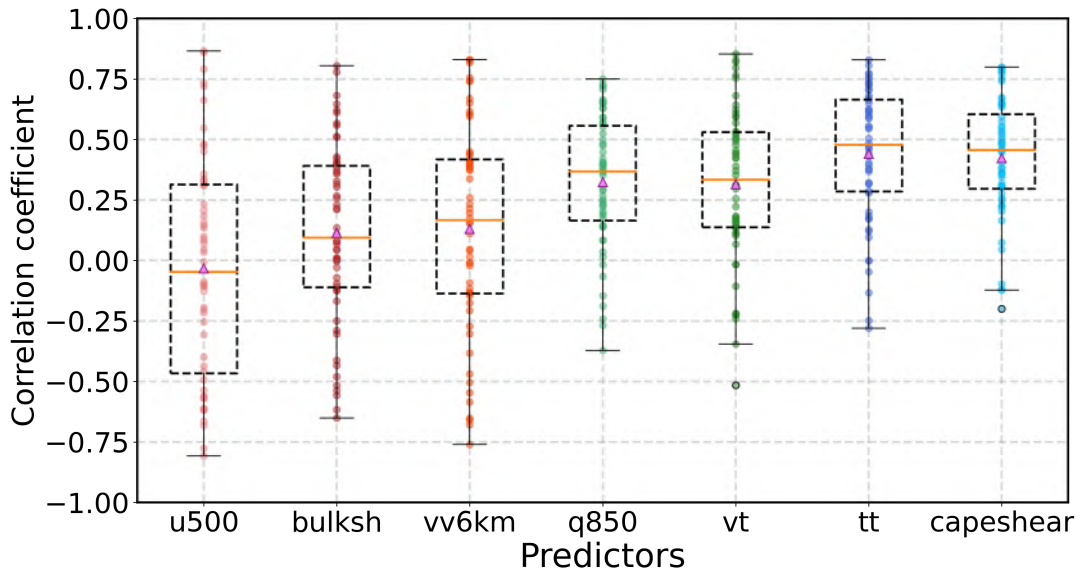
(a) 56 of Correct forecast events
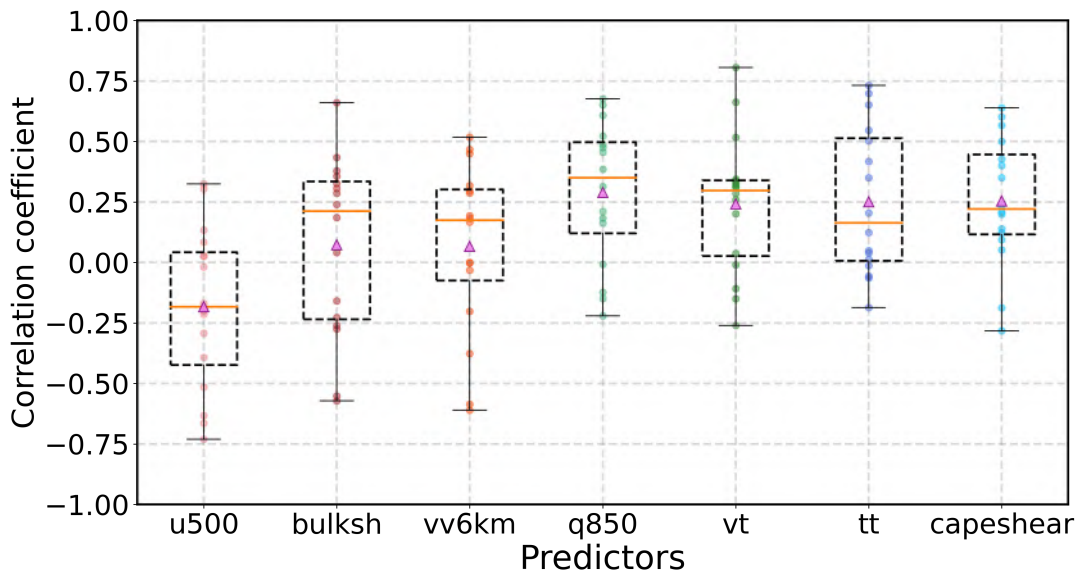
(b) 16 of False alarm events

Figure 6.4: Statistical distributions of the correlation coefficients between heatmaps and predictors in the boxplot, to reflect the importance of each predictor for the ML model identification process. The median is the orange line, and mean is the pink shaded triangle for each predictor to comprehensively characterize the corresponding feature distribution.

whole of Germany could be one approach (Kunz et al., 2020). These may become the next direction of future work.

7 Conclusion and discussion

Predicting severe hail is a complex task for numerical weather prediction (NWP) systems. For the effective hail prediction, these systems can not accurately predict the occurrence, path, and severity of hailstorms. In addition, the representation of hail in these models is closely linked to the microphysical parameterisation, a critical source of uncertainty and potential inaccuracy at the storm level. The rapid progression and changing nature of these storms further limits the predictability of hail, as forecast errors can escalate rapidly at this scale. It's therefore vital to have an accurate understanding of the atmospheric conditions and the inherent uncertainty in these estimates (Snook et al., 2016).

The objective of this study is to investigate the relationship between the preconvective hailstorm environments taken from ERA5 reanalysis data and hailstorm occurrence, represented by radar-derived hailstorm data from 2005 to 2019 during the hail season (April to September) only. More importantly, the overarching goal is to select the most appropriate predictors in order to develop a machine learning model that can accurately predict the hail-affected area by collecting and combining information on convective environmental conditions. In the following, we provide answers to the research questions raised in Chapter 1:

1. Is there a characteristic geographic distribution of the prevailing convective environment for hailstorms in Germany?

Yes, in southern Germany the environment is more conductive to hailstorms than in central and northern Germany, which is reflected in the increasing CAPE, CPARESHEAR and SHIP from north to south. This trend aligns with the fact that the frequency of hailstorms in Germany increases progressively from north to south, an observation which logically corresponds with the gradient in convective conditions, especially the large CAPESHEAR areas overlapping mountainous or sloping terrain in Baden-Württemberg, Bavaria, and also northern Hesse (Punge and Kunz, 2016).

2. Is machine learning useful for hail prediction and for better understanding relationships between ambient conditions and hailstorms? What is the predictive skill of the machine learning models?

Concerning the first question, the findings presented in Chapter 5 confirm that machine learning holds a significant potential to improve hail forecasting. In comparison to traditional, non-machine learning forecasting methods, machine learning exhibits a markedly superior performance, demonstrating the lowest Mean Absolute Errors (MAEs) in terms of model

bias, thus outperforming climatology and persistence forecasts. The machine learning model excels in model evaluation, demonstrating an impressive Heidke skill score (HSS) of 0.66 at a spatial scale of 1740 km^2 for daily affected area by hail, alongside a commendable accuracy rate of 88% when compared to climatology forecast (0.51 HSS) and persistence forecast (0.36 HSS). The model also has a Probability of Detection (POD) of 0.78, a Critical Success Index (CSI) of 0.58, and a False Alarm Ratio (FAR) of 0.3, thus positioning it as a highly competent tool for hail prediction.

3. Which predictors yield the best model performance, and how do they modify the accuracy of the model prediction?

From the 18 pre-selected convective parameter candidates, U500, BULKSH, VV6km, Q850, VT, TT, and CAPESHEAR are selected as the seven best predictive indicators that corresponded to the optimal model results. Notably, CAPESHEAR not only demonstrated the highest correlation with hail occurrence, but also emerged as the best single predictive indicator in the machine learning model. For the parameters representing atmospheric dynamics, wind shear plays an important role for the accuracy of the model prediction. Higher values of VT and TT also enhance the machine learning model's ability to successfully predict hail events, mainly due to the way these parameters accentuated atmospheric vertical instability. The inclusion of humidity parameters improved the ML performance further as they may provide energy from phase change of water vapour for the formation of convective storms. On the other hand, dynamical parameters such as BULKSH and VV6km, primarily influenced on the mid to low-level vertical wind shear, are highly correlated between each other, thereby improving slightly the model performance. While U500, as a representation of the large-scale atmospheric circulation in the middle-upper levels, despite their indirect impact on convective storm formation, might be not as the most effective in providing valuable predictive information for the machine learning model.

4. When and why are machine learning models unable to predict hail occurrence?

When isolated thunderstorms or only a small number of storms occurring on a particular day, from the relative weaker convective environment, specifically extensive regions with high CAPESHEAR values far away to corresponding storm trajectories, may yield erroneous predictions of ML model. Similarly, any occurrence of hailstorms after 12 UTC, derived from radar data with significant time lags, can contribute to incorrect predictions from the machine learning model. This is largely due to the model's inability to access sufficient information about related environmental predictors, consequently inhibiting its predictive accuracy to hail-affected area. When looking at the weather conditions for the hit cases and false alarm cases, U500, which indicates large-scale upper atmospheric dynamics, shows noticeable differences. In hit cases, Germany seemingly experiences stronger westerly winds associated with the northern side of blocking systems or ridges. In contrast, false alarm cases are often associated with weak easterly winds, which may be related to the northern edge of low-pressure systems.

In summary, statistical methods were firstly applied to investigate the distribution of hailstorm-favorable environmental conditions across Germany. The relationship between convective variables and hail-affected area was analyzed to identify potential convective parameters for machine learning, which are for Germany CAPESHEAR, CAPE, VT, TT, SHIP.

Throughout the machine learning training phase, 15 years of ERA5 reanalysis data and radar-identified potential hailstorms were divided into 13 years of training data (2005–2017) and 2 years of test data (2018–2019). The predictive performance of the machine learning model was comprehensively evaluated by comparing its performance separately with that of non-machine learning models and by assessing various skill scores for verification. Machine learning sensitivity experiments were conducted for different times of day, different convective parameters and linear correlation between input data, resulting in the best seven predictors at 12 UTC, which are CAPES-HEAR, q850, VV6km, U500, BULKSH, TT, VT, with the non-linear activation among the input parameters and between different layers of the CNN to keep independence during the training.

Ultimately, the activation heatmap serves as a tool to "open up" the "black box" of machine learning, allowing us to visualize how ML predictions are made. It sheds light on how these seven predictors differently affect and contribute to machine learning's decision-making process. The results indicate that the success of the predictions for the machine learning model is related to the clusters of hailstorms and the stronger convective environmental field signals that occurred on the day. Machine learning still cannot accurately predict and localize the occurrence of individual storms or more dispersed storms with relative weak prevailing convective conditions around.

The performance of machine learning models in predicting the hail-affected area on a daily basis demonstrated in this study is promising. However, the existing uncertainty of the input data in combination with the sole use of machine learning predictions to forecast severe convective storms remains a puzzle. Therefore, the accuracy of the prediction necessitates reinforcement from other data sources through other predictive model formats, such as ensemble forecasting. On another front, the key to optimising machine learning models is to obtain as much information as possible about the variables associated with hailstorm environments, and in addition to the idea of collecting convective conditions of the center of hail tracks mentioned in Chapter 6, the temporal evolution of the predictors may also provide opportunities for machine learning optimisation. Furthermore, machine learning models could potentially provide additional information on prevailing convective condition otherwise unattainable through numerical forecasts, for example,it may play an important role in the physical parameterization for model simulation, thereby fine-tuning the accuracy of numerical model predictions. This unique attribute can potentially underscores the critical role of machine learning in improving the accuracy of weather forecast.
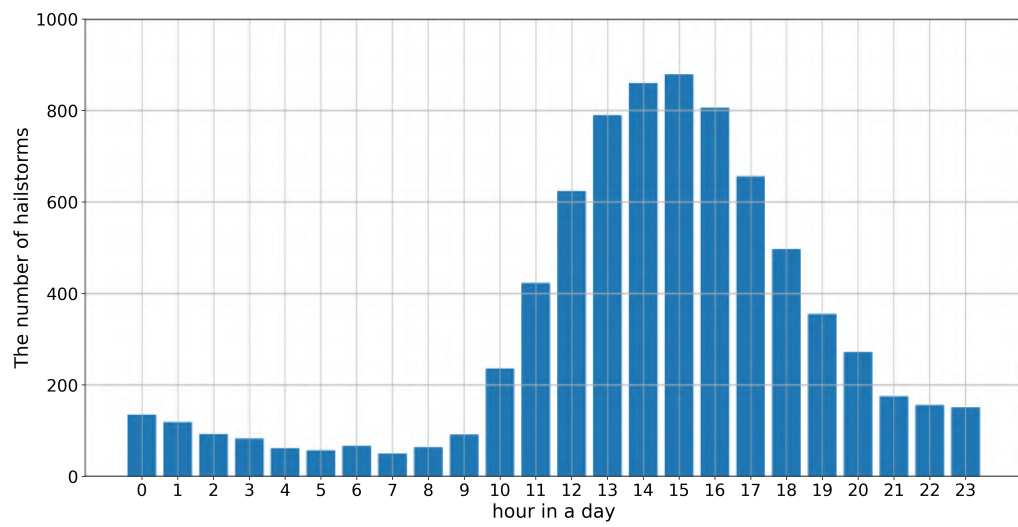
Appendix

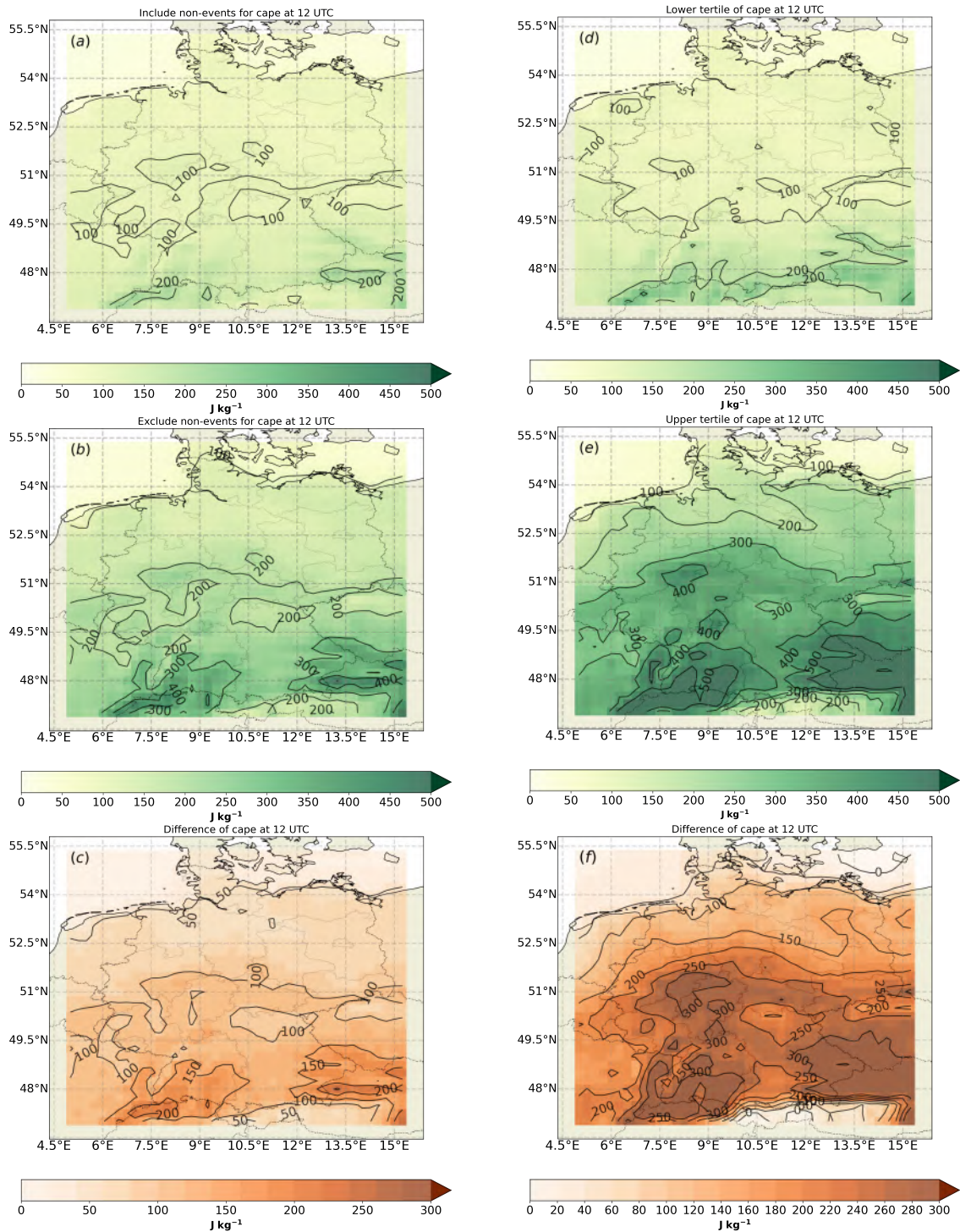Figure A.1: Histogram of hailstorms in daily time (UTC).

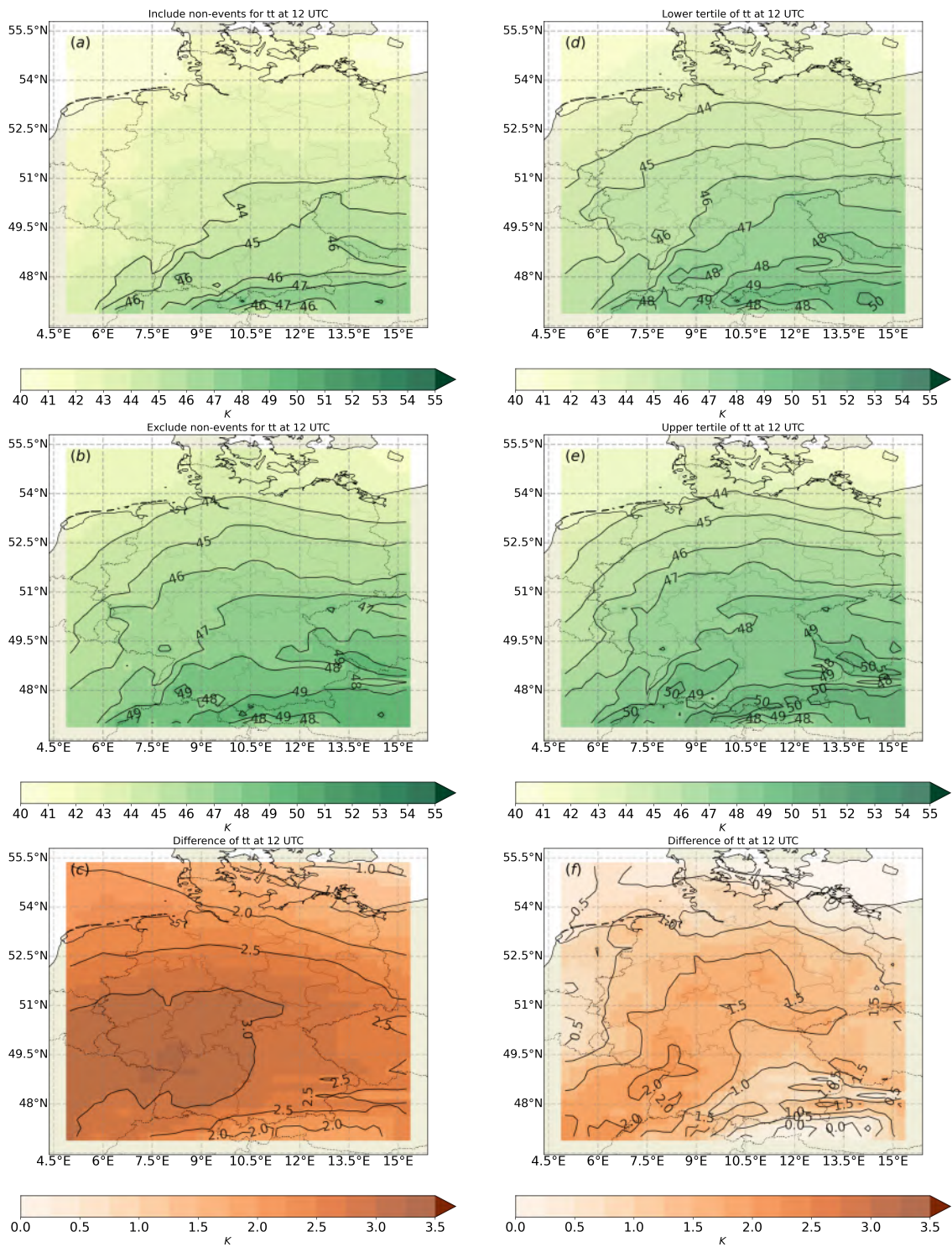Figure A.2: Mean fields of CAPE during the hail season from 2005-2019. Same as Figure 4.4.

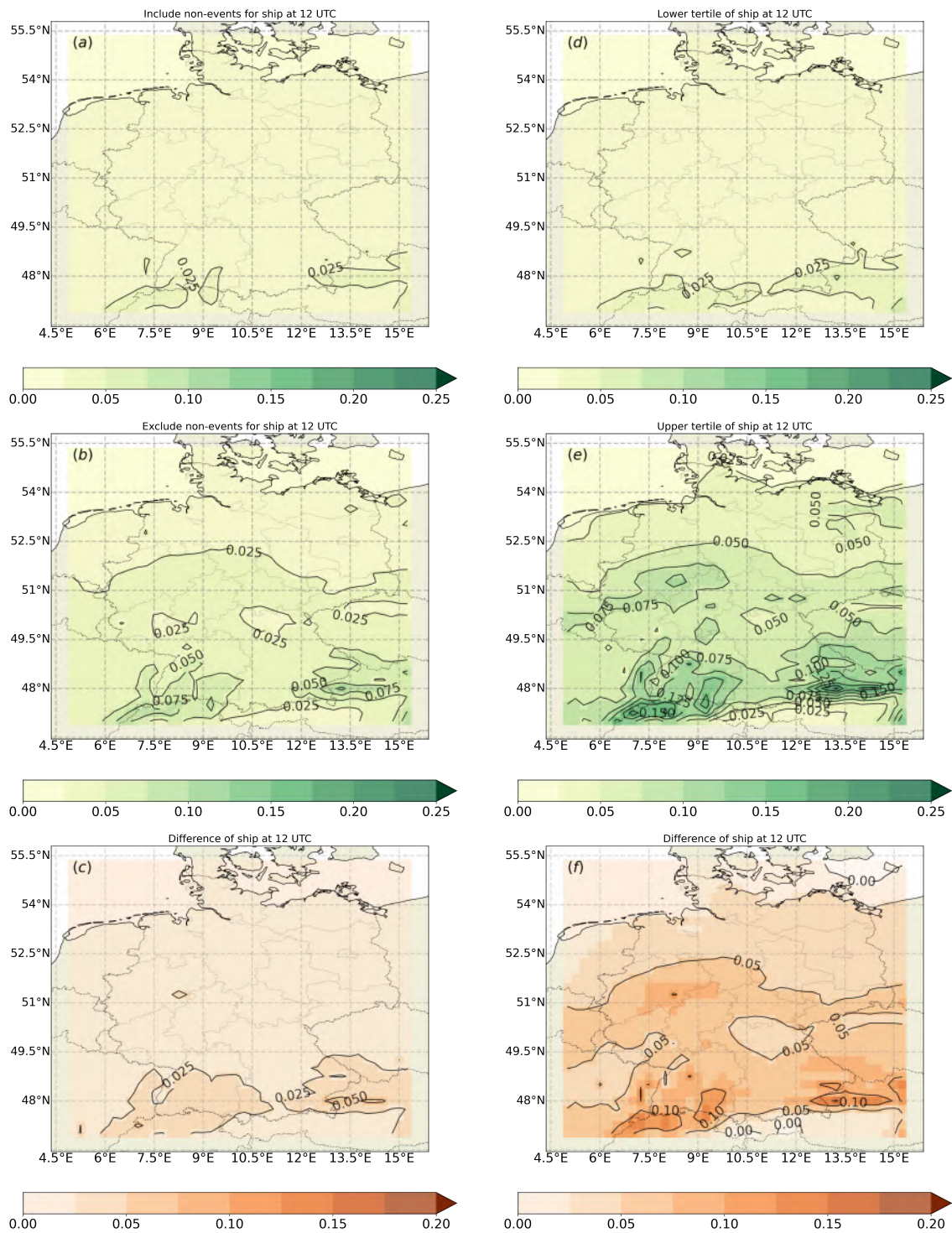Figure A.3: Mean fields of TT during the hail season from 2005-2019. Same as Figure 4.4.

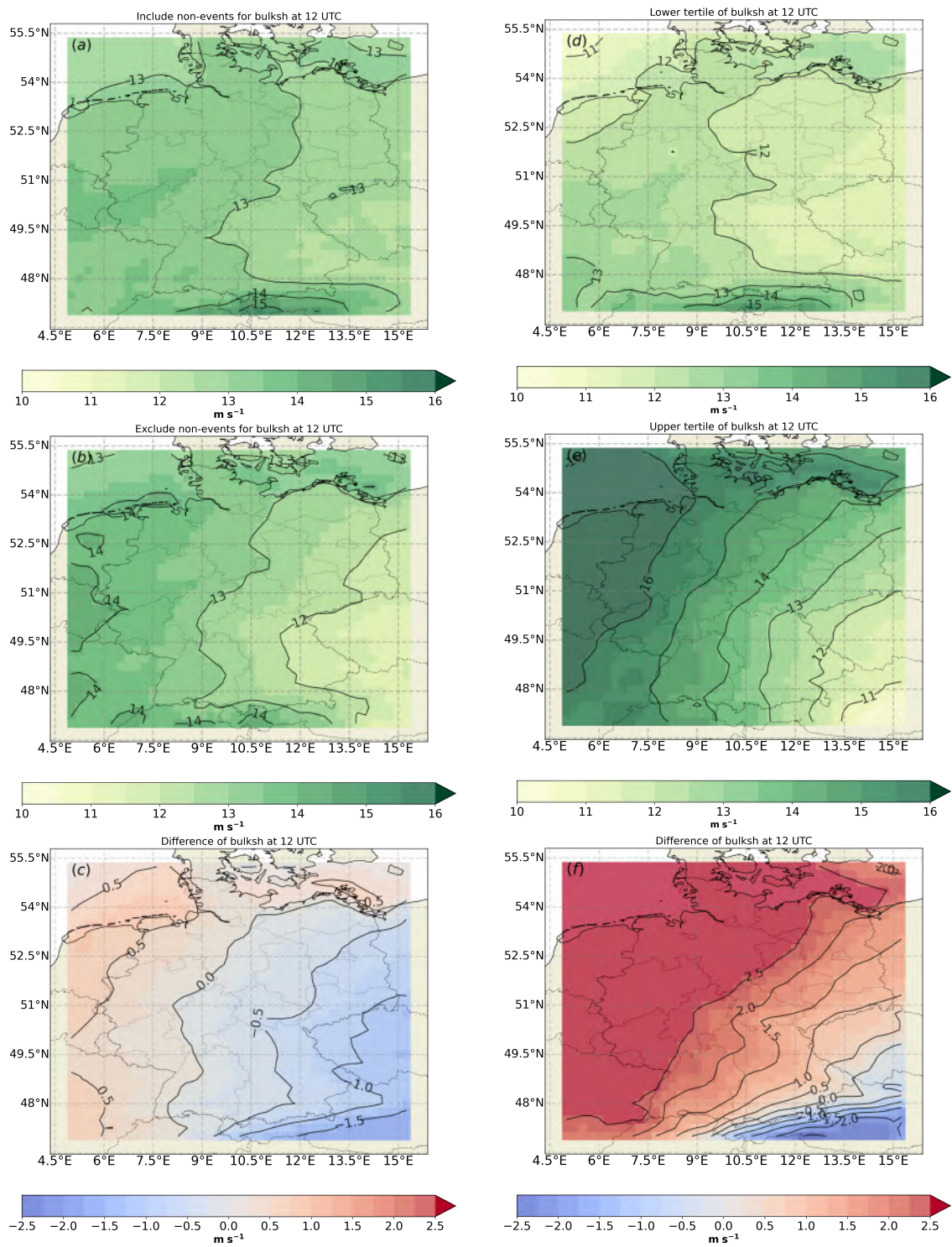Figure A.4: Mean fields of SHIP during the hail season from 2005-2019. Same as Figure 4.4.

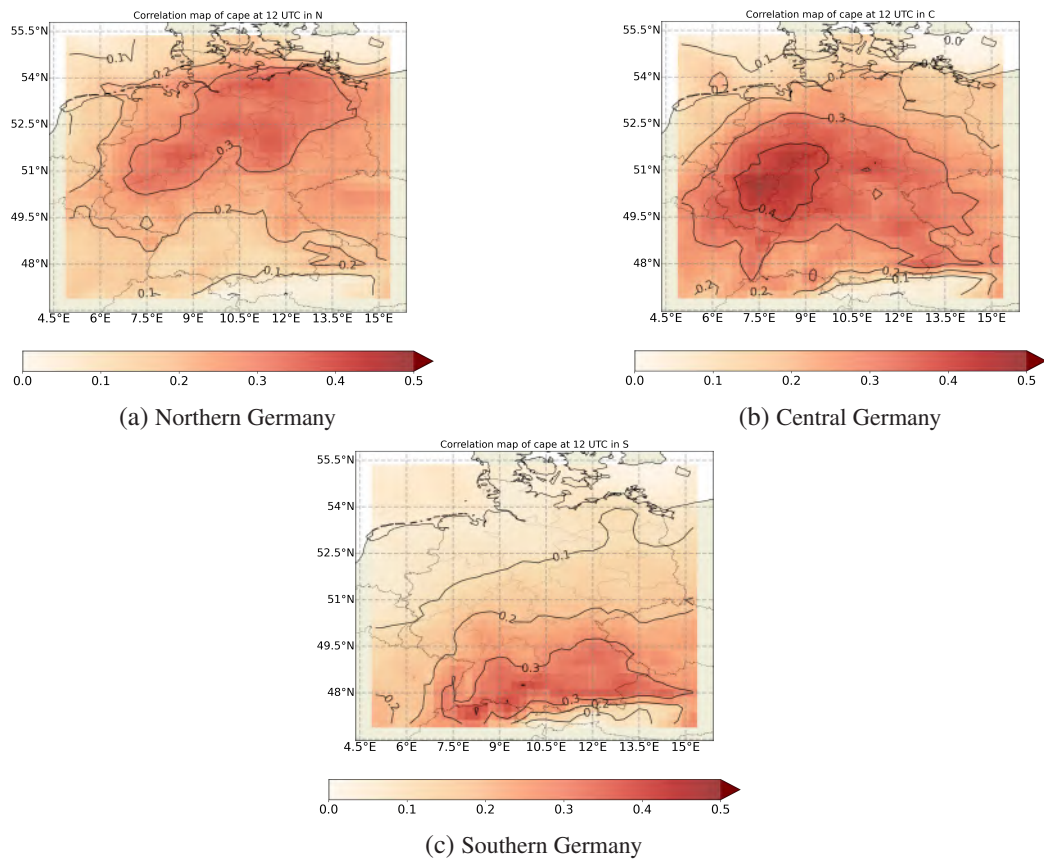Figure A.5: Mean fields of BULKSH during the hail season from 2005-2019. Same as Figure 4.4.

(a) Northern Germany

(b) Central Germany

(c) Southern Germany

Figure A.6: Correlation maps between CAPE and the daily affected area, same as Figure 4.5.

(a) Northern Germany

(b) Central Germany

(c) Southern Germany
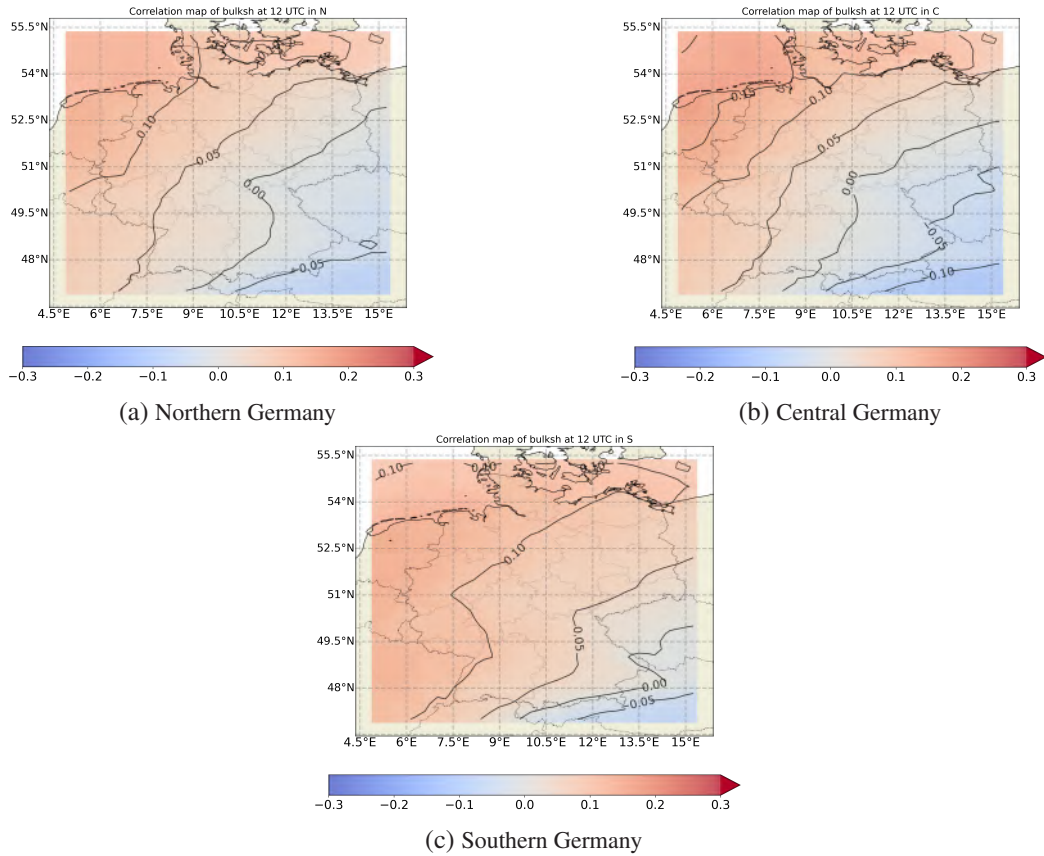
Figure A.7: Correlation maps between BULKSH and the daily affected area, same as Figure 4.5.

(a) Northern Germany

(b) Central Germany

(c) Southern Germany
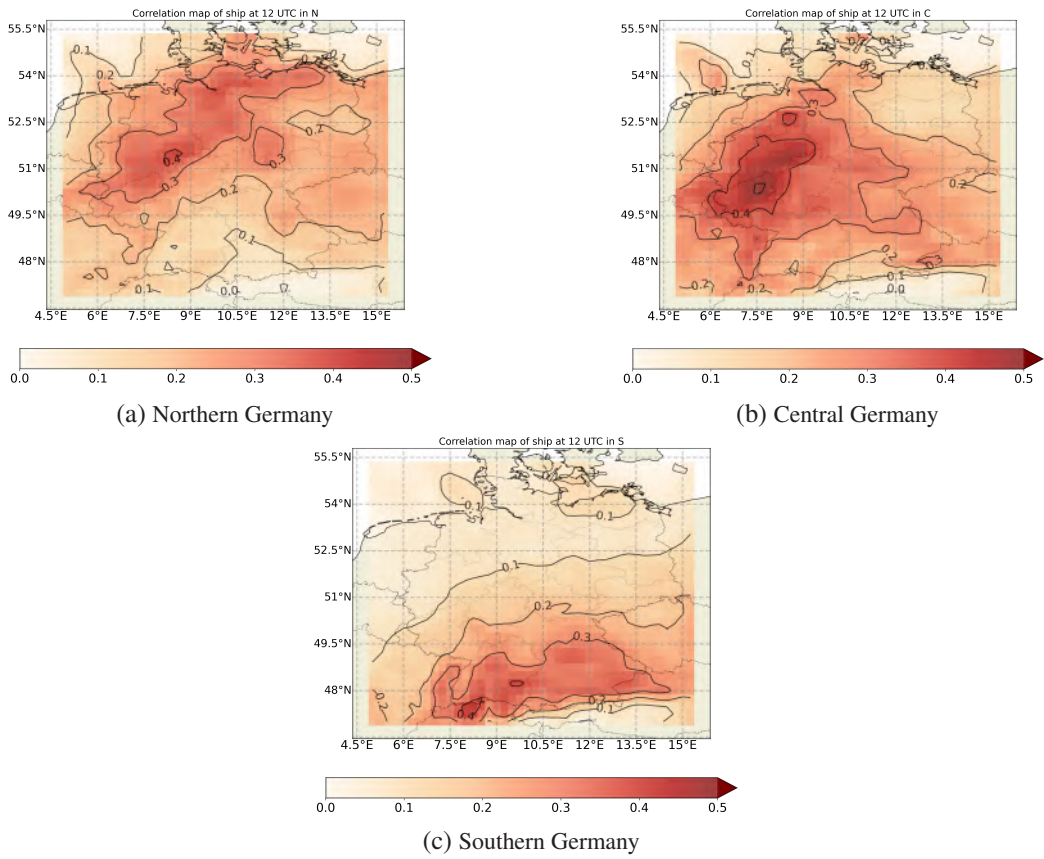
Figure A.8: Correlation maps between SHIP and the daily affected area, same as Figure 4.5.

(a) Northern Germany

(b) Central Germany

(c) Southern Germany
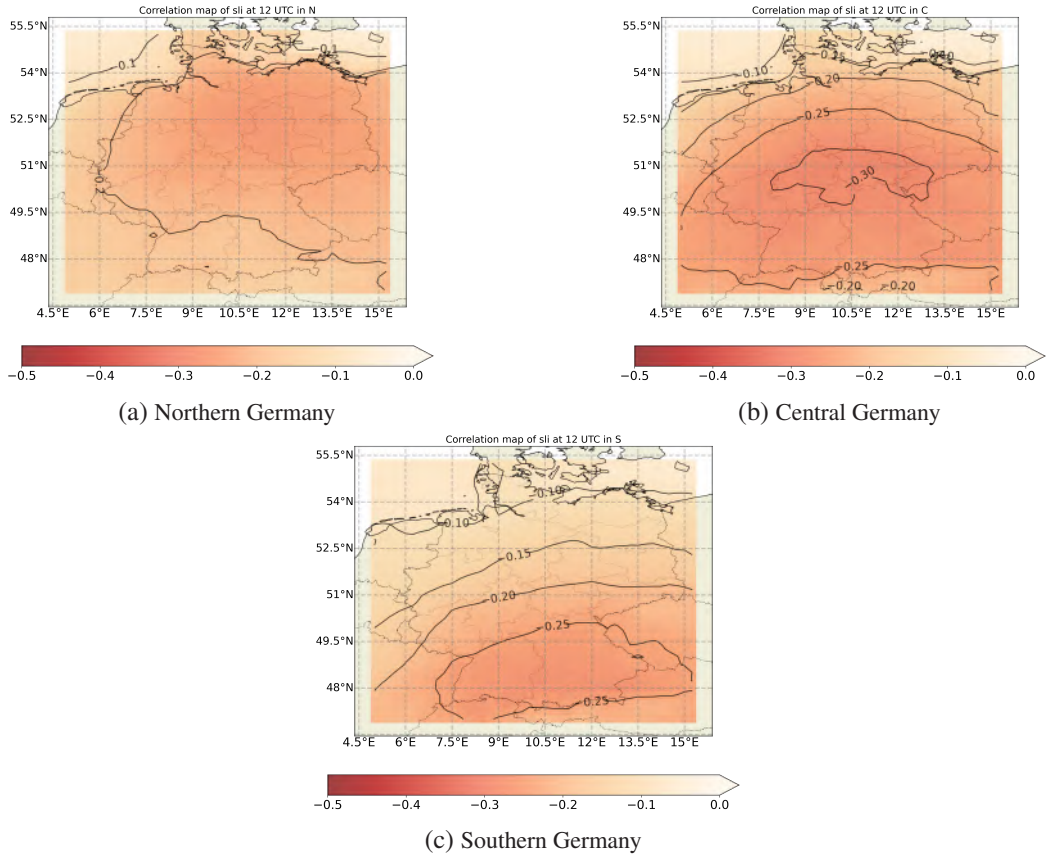
Figure A.9: Correlation maps between SLI and the daily affected area, same as Figure 4.5.

(a)
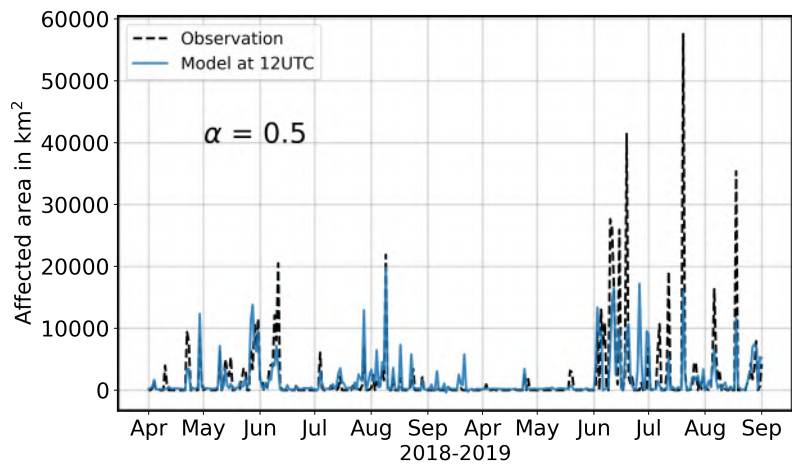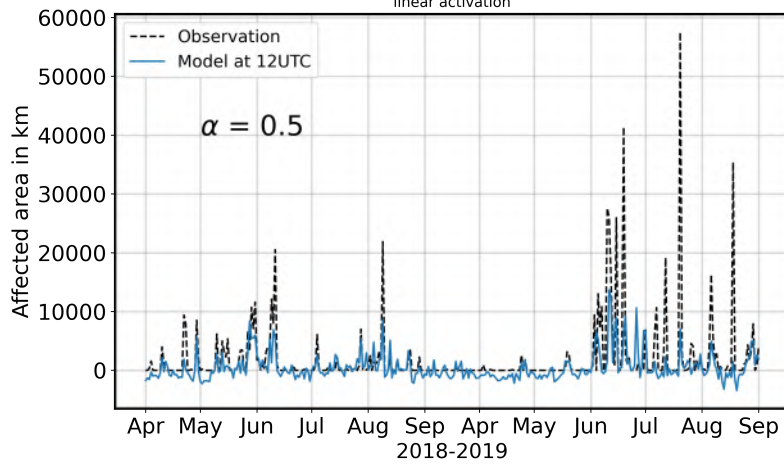
(b)
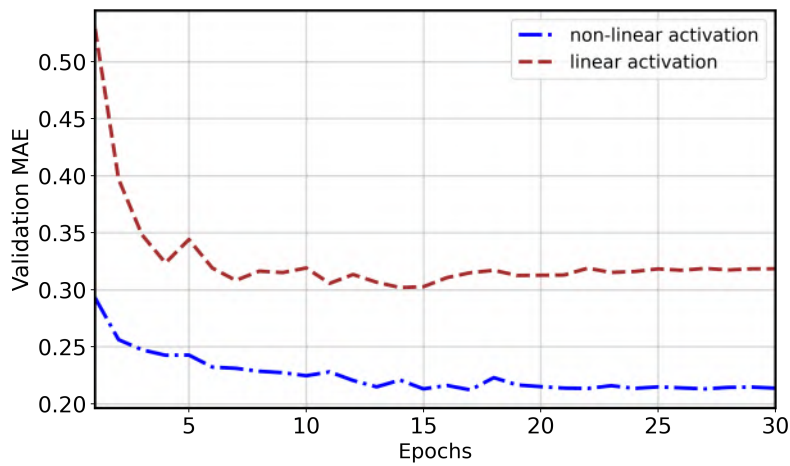
(c)

Figure A.10: The outputs of the convolutional neural network at 12 UTC. (a) are the affected area predicted for non-linear activation function in CNNs, (b) is for linear activation, both with 0.5 of α. The blue solid line is the ML prediction and the black dashed line is the observation for comparison. (c) is the validation MAEs of CNN output at 12UTC, as a function of training epochsof model.
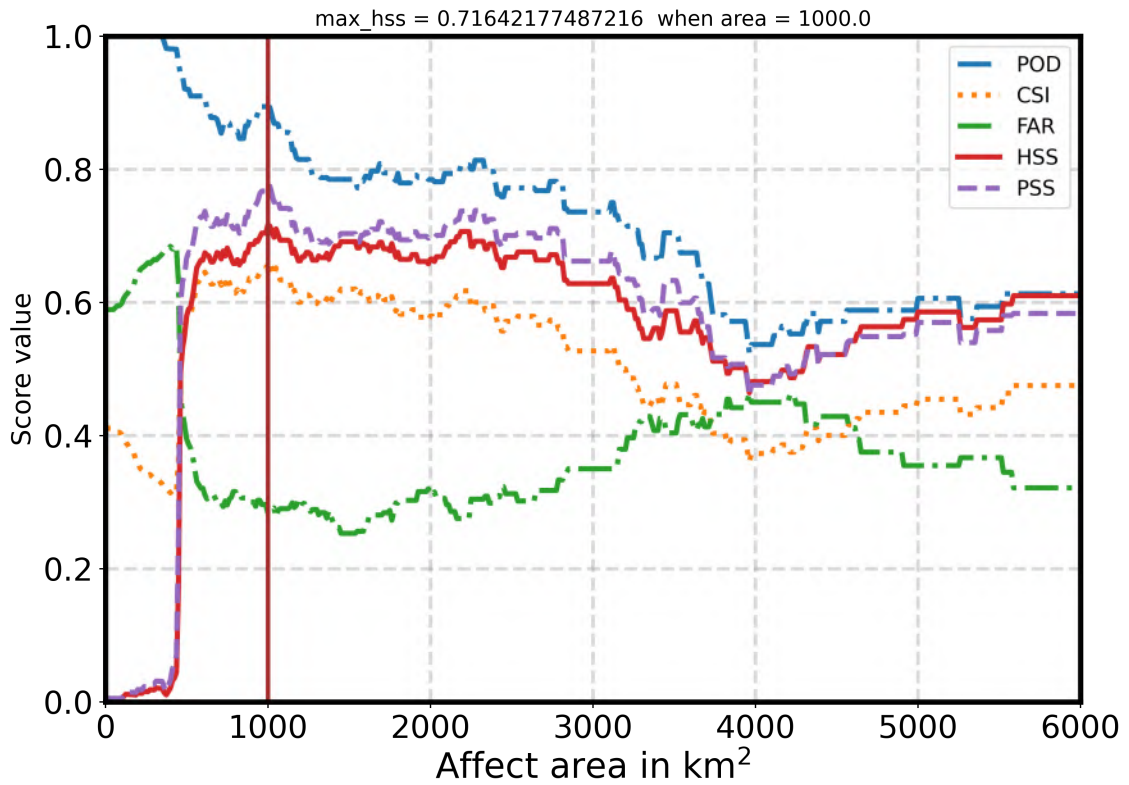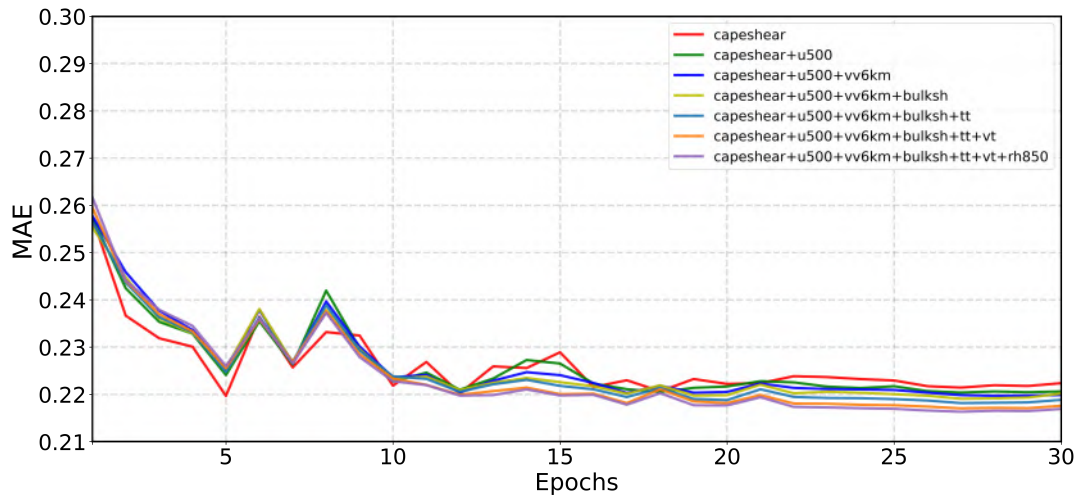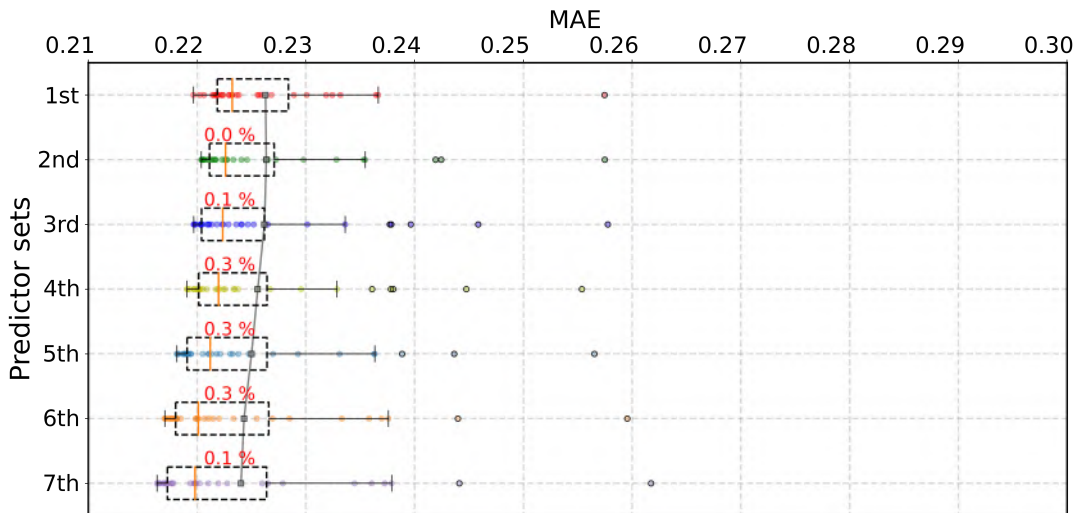
Figure A.11: Same as Figure 5.3, but at 18 UTC.
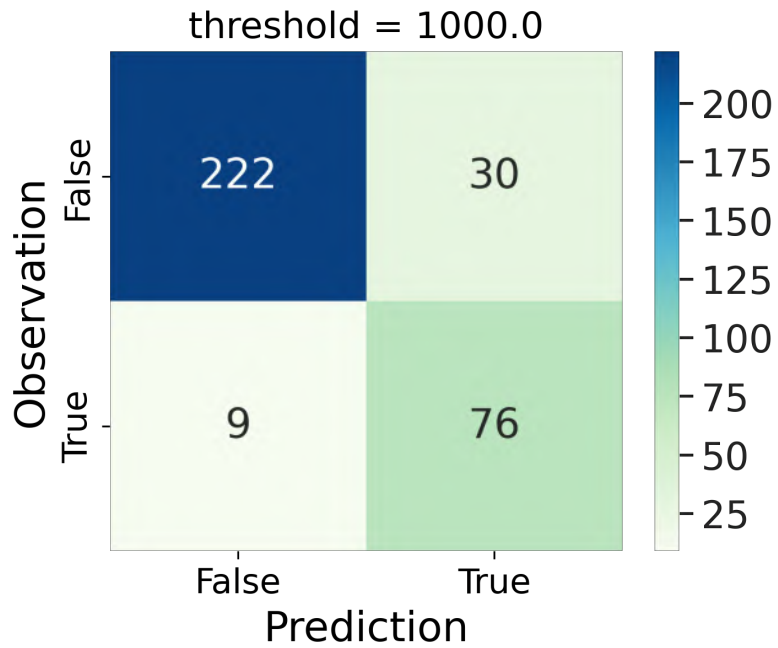
Figure A.12: Contingency table at the optimal threshold of 1000 km^2 for the daily affected area at 18 UTC.

(a) The variation of MAEs as a function of the number of iterations during training for the sensitivity test.

(b) Boxplot of the MAEs distribution in each training period. The Grey line is the variation of averaged MAE of each training. Percentage numbers represent the relative increased rate of MAEs compared with the last training.

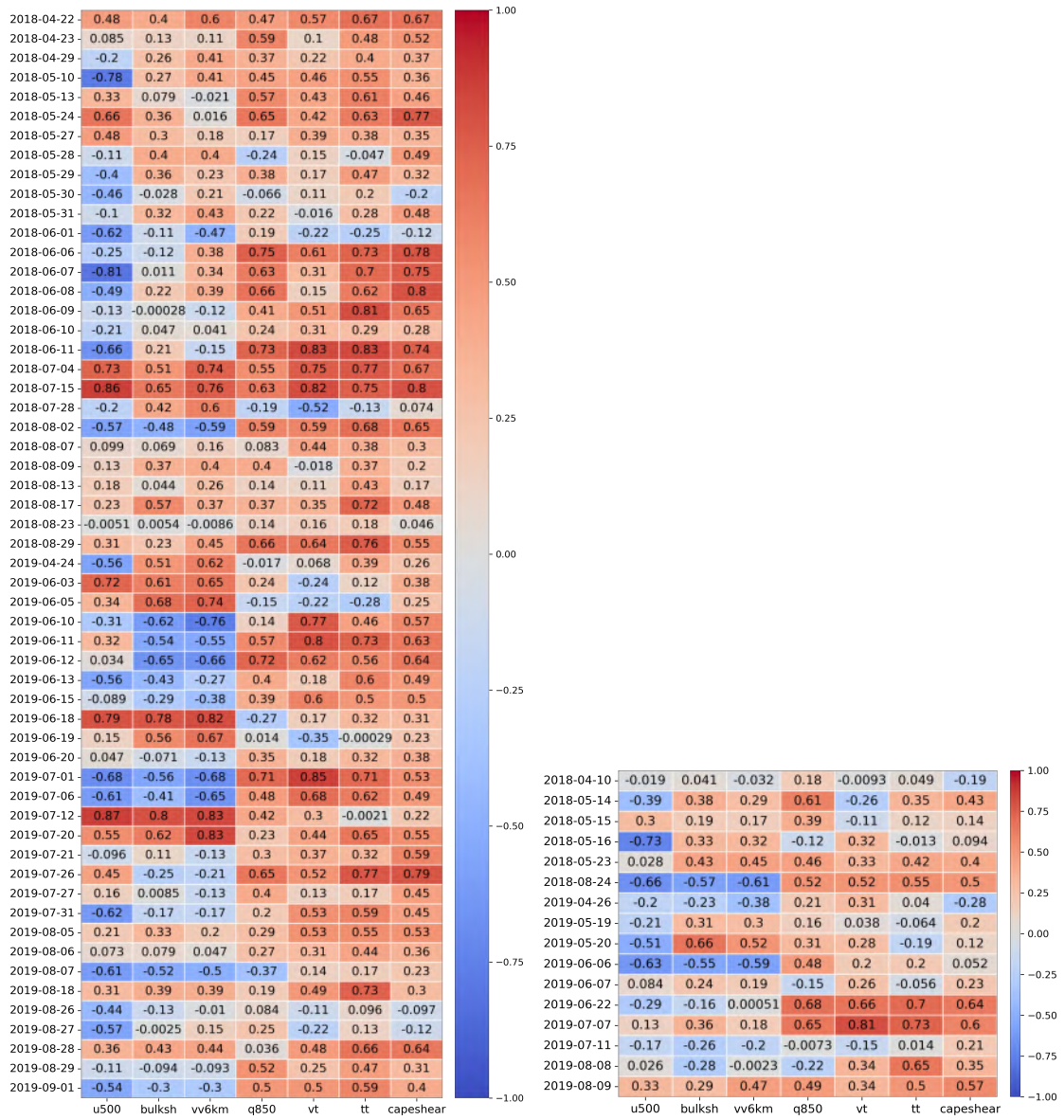Figure A.13: Model performance when preconditions at 18 UTC, same as the Figure 6.1

(a) For 56 correct forecast events in Figure 5.4.

(b) For 16 false alarm events in Figure 5.4.

Figure A.14: Correlation matrix between ambient predictors and the activation heatmaps of 72 observed hail events at 12 UTC for the preconditions.

Bibliography

Ali, P. J. M., R. H. Faraj, E. Koya, P. J. M. Ali, and R. H. Faraj, 2014: Data normalization and standardization: a technical report. *Mach Learn Tech Rep*, **1 (1)**, 1–6.

Aran, M., J. Pena, and M. Torà, 2011: Atmospheric circulation patterns associated with hail events in lleida (catalonia). *Atmospheric research*, **100 (4)**, 428–438.

Atlas, D., 2015: *Radar in Meteorology: Battan Memorial and 40th Anniversary Radar Meteorology Conference*. Springer.

Auer, A. H., 1972: Distribution of graupel and hail with size. *Monthly Weather Review*, **100 (5)**, 325–328.

Bartels, H., and Coauthors, 2005: Projekt radvor-op: Radargestützte, zeitnahe niederschlagsvorher-sage für den operationellen einsatz (niederschlag-nowcasting-system), final report, deutscher wetterdienst. *Final Report, Deutscher Wetterdienst available via: http://www. dwd. de/radvor-op*.

Bean, B. R., and E. Dutton, 1966: *Radio meteorology*, Vol. 92. Superintendentof Documents, US GovernmentPrint. Office.

Bedka, K. M., 2011: Overshooting cloud top detections using msg seviri infrared brightness temperatures and their relationship to severe weather over europe. *Atmospheric Research*, **99 (2)**, 175–189.

Billingsley, J. B., 1993: Ground clutter measurements for surface-sited radar. Tech. rep., MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB.

Boddy, R., and G. Smith, 2009: *Statistical methods in practice: for scientists and technologists*. John Wiley & Sons.

Bolton, D., 1980: The computation of equivalent potential temperature. *Monthly weather review*, **108 (7)**, 1046–1053.

Brooks, H. E., 2009: Proximity soundings for severe convection for europe and the united states from reanalysis data. *Atmospheric Research*, **93 (1-3)**, 546–553.

Brooks, H. E., 2013: Severe thunderstorms and climate change. *Atmospheric research*, **123**, 129–138.

Brown, L., 1999: A radar history of world war ii. *J. Am. Hist. Res*.

Byers, H. R., and R. R. Braham, 1949: *The thunderstorm: report of the Thunderstorm Project*. US Government Printing Office.

Cao, Z., 2008: Severe hail frequency over ontario, canada: Recent trend and variability. *Geophysical Research Letters*, **35 (14)**.

Chai, T., and R. R. Draxler, 2014a: Root mean square error (rmse) or mean absolute error (mae). *Geoscientific model development discussions*, **7 (1)**, 1525–1534.

Chai, T., and R. R. Draxler, 2014b: Root mean square error (rmse) or mean absolute error (mae)?– arguments against avoiding rmse in the literature. *Geoscientific model development*, **7 (3)**, 1247–1250.

Changnon, S. A., 1977: The scales of hail. *Journal of Applied Meteorology and Climatology*, **16 (6)**, 626–648.

Changnon Jr, S. A., 1970: Hailstreaks. *Journal of Atmospheric Sciences*, **27 (1)**, 109–125.

Chen, S. S., and W. M. Frank, 1993: A numerical study of the genesis of extratropical convective mesovortices. part i: Evolution and dynamics. *Journal of the atmospheric sciences*, **50 (15)**, 2401–2426.

Chollet, F., 2021: *Deep learning with Python*. Simon and Schuster.

Chỳlek, P., 1976: Partial-wave resonances and the ripple structure in the mie normalized extinction cross section. *JOSA*, **66 (3)**, 285–287.

Cohen, I., Y. Huang, J. Chen, J. Benesty, J. Benesty, J. Chen, Y. Huang, and I. Cohen, 2009: Pearson correlation coefficient. *Noise reduction in speech processing*, 1–4.

Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Monthly Weather Review*, **124 (8)**, 1767–1785.

Czernecki, B., M. Taszarek, M. Marosz, M. Półrolniczak, L. Kolendowicz, A. Wyszogrodzki, and J. Szturc, 2019: Application of machine learning to large hail prediction-the importance of radar reflectivity, lightning occurrence and convective parameters derived from era5. *Atmospheric Research*, **227**, 249–262.

Diffenbaugh, N. S., M. Scherer, and R. J. Trapp, 2013: Robust increases in severe thunderstorm environments in response to greenhouse forcing. *Proceedings of the National Academy of Sciences*, **110 (41)**, 16 361–16 366.

Dixon, M., and G. Wiener, 1993: Titan: Thunderstorm identification, tracking, analysis, and nowcasting—a radar-based methodology. *Journal of atmospheric and oceanic technology*, **10 (6)**, 785–797.

Doswell, C., R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Weather and forecasting*, **5 (4)**, 576–585.

Doswell, C. A., 1985: The operational meteorology of convective weather. volume ii. storm scale analysis.

Doswell III, C. A., 1987: The distinction between large-scale and mesoscale contribution to severe convection: A case study example. *Weather and Forecasting*, **2 (1)**, 3–16.

Ebert, E., and J. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *Journal of hydrology*, **239 (1-4)**, 179–202.

ECMWF, 2010a: Parameter details - total column water. URL https://codes.ecmwf.int/grib/param-db/?id=136.

ECMWF, 2010b: Parameter details - total totals index. URL https://codes.ecmwf.int/grib/param-db/?id=260123.

ECMWF, 2022: What are the changes from era-interim to era5 and era5-land? URL :https://confluence.ecmwf.int/pages/viewpage.action?pageId=74764925.

ECMWF, 2023: Era5: data documentation. URL :https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#heading-Relatedarticles.

Fluck, E., 2018: Hail statistics for european countries. Ph.D. thesis, KIT-Bibliothek.

Fluck, E., M. Kunz, P. Geissbuehler, and S. P. Ritz, 2021: Radar-based assessment of hail frequency in europe. *Natural Hazards and Earth System Sciences*, **21 (2)**, 683–701.

Gagne, D., A. McGovern, J. Jerald, M. Coniglio, J. Correia, and M. Xue, 2015: Day-ahead hail prediction integrating machine learning with storm-scale numerical weather models. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29, 3954–3960.

Gagne, D. J., A. McGovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Weather and forecasting*, **32 (5)**, 1819–1840.

Gagne II, D. J., S. E. Haupt, D. W. Nychka, and G. Thompson, 2019: Interpretable deep learning for spatial analysis of severe hailstorms. *Monthly Weather Review*, **147 (8)**, 2827–2845.

Galway, J. G., 1956: The lifted index as a predictor of latent instability. *Bulletin of the American Meteorological Society*, **37 (10)**, 528–529.

Gensini, V. A., and T. L. Mote, 2015: Downscaled estimates of late 21st century severe weather from ccsm3. *Climatic Change*, **129 (1)**, 307–321.

Giaiotti, D., S. Nordio, and F. Stel, 2003: The climatology of hail in the plain of friuli venezia giulia. *Atmospheric Research*, **67**, 247–259.

Grieser, J., 2012: *Convection parameters*. Selbstverl.

Gudd, M., 2004: Gewitter und gewitterschäden im südlichen hessischen berg-und beckenland und im rhein-main-tiefland 1881 bis 1980. Ph.D. thesis, Johannes Gutenberg-Universität Mainz.

Gunn, K. L. S., and T. W. R. East, 1954: The microwave properties of precipitation particles. *Quarterly Journal of the Royal Meteorological Society*, **80 (346)**, 522–545.

Handwerker, J., 2002: Cell tracking with trace3d—a new algorithm. *Atmospheric Research*, **61 (1)**, 15–34.

Hauke, J., and T. Kossowski, 2011: Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, **30 (2)**, 87–93.

Haurwitz, B., 1941: Dynamic meteorology. Tech. rep.

Heidke, P., 1926: Berechnung des erfolges und der güte der windstärkevorhersagen im sturmwarnungsdienst. *Geografiska annaler*, **8 (4)**, 301–349.

Helmert, K., and Coauthors, 2014: Dwds new radar network and post-processing algorithm chain. *Proc. Eighth European Conf. on Radar in Meteorology and Hydrology (ERAD 2014), Garmisch-Partenkirchen, Germany, DWD and DLR*, Vol. 4, 1–5.

Hersbach, H., and Coauthors, 2020: The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146 (730)**, 1999–2049.

Holleman, I., H. Wessels, J. Onvlee, and S. Barlag, 2000: Development of a hail-detection-product: S10: Deep convection. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, **25 (10-12)**, 1293–1297.

Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society*, **79 (9)**, 1855–1870.

Hughes, P., 1983: A high-resolution radar detection strategy. *IEEE Transactions on Aerospace and Electronic Systems*, **(5)**, 663–667.

Jones, A., 1979: Electromagnetic wave scattering by assemblies of particles in the rayleigh approximation. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, **366 (1724)**, 111–127.

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

Krizhevsky, A., I. Sutskever, G. E. Hinton, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, 2012: Advances in neural information processing systems. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United . . . .

Kunz, M., 2007: The skill of convective parameters and indices to predict isolated and severe thunderstorms. *Natural Hazards and Earth System Sciences*, **7 (2)**, 327–342, https://doi.org/10.5194/nhess-7-327-2007, URL https://nhess.copernicus.org/articles/7/327/2007/.

Kunz, M., U. Blahak, J. Handwerker, M. Schmidberger, H. J. Punge, S. Mohr, E. Fluck, and K. M. Bedka, 2018: The severe hailstorm in southwest germany on 28 july 2013: Characteristics,

impacts and meteorological conditions. *Quarterly Journal of the Royal Meteorological Society*, **144 (710)**, 231–250.

Kunz, M., and P. I. Kugel, 2015: Detection of hail signatures from single-polarization c-band radar reflectivity. *Atmospheric Research*, **153**, 565–577.

Kunz, M., and M. Puskeiler, 2010: High-resolution assessment of the hail hazard over complex terrain from radar and insurance data. *Meteorologische Zeitschrift*, **19 (5)**, 427.

Kunz, M., J. Sander, and C. Kottmeier, 2009: Recent trends of thunderstorm and hailstorm frequency and their relation to atmospheric characteristics in southwest germany. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, **29 (15)**, 2283–2297.

Kunz, M., J. Wandel, E. Fluck, S. Baumstark, S. Mohr, and S. Schemm, 2020: Ambient conditions prevailing during hail events in central europe. *Natural Hazards and Earth System Sciences*, **20 (6)**, 1867–1887.

Lamb, D., and J. Verlinde, 2011: *Physics and chemistry of clouds*. Cambridge University Press.

Lee, J. W., 2002: Tornado proximity soundings from the ncep/ncar reanalysis data.

Lemon, L. R., and C. A. Doswell III, 1979: Severe thunderstorm evolution and mesocyclone structure as related to tornadogenesis. *Monthly Weather Review*, **107 (9)**, 1184–1197.

Maddox, R. A., K. W. Howard, D. L. Bartels, and D. M. Rodgers, 1986: Mesoscale convective complexes in the middle latitudes. *Mesoscale meteorology and forecasting*, 390–413.

Markowski, P., and Y. Richardson, 2011: *Mesoscale meteorology in midlatitudes*, Vol. 2. John Wiley & Sons.

Marshall, J., W. Hitschfeld, and K. Gunn, 1955: Advances in radar weather. *Advances in geophysics*, Vol. 2, Elsevier, 1–56.

Marwitz, J. D., 1972: The structure and motion of severe hailstorms. part ii: Multi-cell storms. *Journal of Applied Meteorology (1962-1982)*, 180–188.

Mason, B., 1971: The physics of clouds, clarendon. Oxford.

Mohr, S., and M. Kunz, 2013: Recent trends and variabilities of convective parameters relevant for hail events in germany and europe. *Atmospheric Research*, **123**, 211–228.

Mohr, S., M. Kunz, and K. Keuler, 2015: Development and application of a logistic model to estimate the past and future hail potential in germany. *Journal of Geophysical Research: Atmospheres*, **120 (9)**, 3939–3956.

Mohr, S., J. Wandel, S. Lenggenhager, and O. Martius, 2019: Relationship between atmospheric blocking and warm-season thunderstorms over western and central europe. *Quarterly Journal of the Royal Meteorological Society*, **145 (724)**, 3040–3056.

Mohr, S., and Coauthors, 2020: The role of large-scale dynamics in an exceptional sequence of severe thunderstorms in europe may–june 2018. *Weather and Climate Dynamics*, **1 (2)**, 325–348.

Moncrieff, M. W., and M. J. Miller, 1976: The dynamics and simulation of tropical cumulonimbus and squall lines. *Quarterly Journal of the Royal Meteorological Society*, **102 (432)**, 373–394.

MunichRE, 2020: URL https://www.munichre.com/de/unternehmen/media-relations/ medieninformationen-und-unternehmensnachrichten/medieninformationen/2020/ milliardenschaeden-praegen-bilanz-naturkatastrophen-2019.html.

Murphy, A. H., 1992: Climatology, persistence, and their linear combination as standards of reference in skill scores. *Weather and forecasting*, **7 (4)**, 692–698.

Murphy, A. H., 1996: The finley affair: A signal event in the history of forecast verification. *Weather and forecasting*, **11 (1)**, 3–20.

Parker, M., 2010: Chapter 16 - radar basics. *Digital Signal Processing 101*, M. Parker, Ed., Newnes, Boston, 191–200, https://doi.org/https://doi.org/10.1016/B978-1-85617-921-8.00020-1, URL https://www.sciencedirect.com/science/article/pii/B9781856179218000201.

Pasini, A., P. Racca, S. Amendola, G. Cartocci, and C. Cassardo, 2017: Attribution of recent temperature behaviour reassessed by a neural-network method. *Scientific reports*, **7 (1)**, 17 681.

Pedamonti, D., 2018: Comparison of non-linear activation functions for deep neural networks on mnist classification task. *arXiv preprint arXiv:1804.02763*.

Piper, D., and M. Kunz, 2017: Spatiotemporal variability of lightning activity in europe and the relation to the north atlantic oscillation teleconnection pattern. *Natural Hazards and Earth System Sciences*, **17 (8)**, 1319–1336.

Prein, A. F., and G. J. Holland, 2018: Global estimates of damaging hail hazard. *Weather and Climate Extremes*, **22**, 10–23.

Pruppacher, H. R., J. D. Klett, and P. K. Wang, 1998: Microphysics of clouds and precipitation. Taylor & Francis.

Púčik, T., P. Groenemeijer, D. Rýva, and M. Kolář, 2015: Proximity soundings of severe and nonsevere thunderstorms in central europe. *Monthly Weather Review*, **143 (12)**, 4805–4821.

Punge, H. J., and M. Kunz, 2016: Hail observations and hailstorm characteristics in europe: A review. *Atmospheric Research*, **176**, 159–184.

Puskeiler, M., M. Kunz, and M. Schmidberger, 2016: Hail statistics for germany derived from single-polarization radar data. *Atmospheric Research*, **178**, 459–470.

Ren, X., X. Li, K. Ren, J. Song, Z. Xu, K. Deng, and X. Wang, 2021: Deep learning-based weather prediction: a survey. *Big Data Research*, **23**, 100 178.

Sauvageot, H., 1992: *Radar meteorology*. Artech House Radar Library (Ha).

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Weather and forecasting*, **5 (4)**, 570–575.

Scher, S., and G. Messori, 2018: Predicting weather forecast uncertainty with machine learning. *Quarterly Journal of the Royal Meteorological Society*, **144 (717)**, 2830–2841.

Schmidberger, M., 2018: *Hagelgefährdung und Hagelrisiko in Deutschland basierend auf einer Kombination von Radardaten und Versicherungsdaten*, Vol. 78. KIT Scientific Publishing.

Seltmann, J. E., T. Hohmann, M. Frech, and P. Tracksdorf, 2013: Dwd's new operational scan strategy. *36rd AMS Conf. on Radar Meteorology, Breckenridge, CO*.

Selvaraju, R. R., M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, 2017: Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE international conference on computer vision*, 618–626.

Seneviratne, S., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment.

Skolnik, M. I., 1962: Introduction to radar. *Radar handbook*, **2**, 21.

Snook, N., Y. Jung, J. Brotzge, B. Putnam, and M. Xue, 2016: Prediction and ensemble forecast verification of hail in the supercell storms of 20 may 2013. *Weather and Forecasting*, **31 (3)**, 811–825.

Steininger, M., K. Kobs, P. Davidson, A. Krause, and A. Hotho, 2021: Density-based weighting for imbalanced regression. *Machine Learning*, **110**, 2187–2211.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the rapid update cycle. *Weather and Forecasting*, **18 (6)**, 1243–1261.

Trapp, R. J., N. S. Diffenbaugh, H. E. Brooks, M. E. Baldwin, E. D. Robinson, and J. S. Pal, 2007a: Changes in severe thunderstorm environment frequency during the 21st century caused by anthropogenically enhanced global radiative forcing. *Proceedings of the National Academy of Sciences*, **104 (50)**, 19 719–19 723.

Trapp, R. J., B. A. Halvorson, and N. S. Diffenbaugh, 2007b: Telescoping, multimodel approaches to evaluate extreme convective weather under future climates. *Journal of Geophysical Research: Atmospheres*, **112 (D20)**.

Trapp, R. J., E. D. Robinson, M. E. Baldwin, N. S. Diffenbaugh, and B. R. Schwedler, 2011: Regional climate of hazardous convective weather through high-resolution dynamical downscaling. *Climate dynamics*, **37 (3)**, 677–688.

Ukkonen, P., A. Manzato, and A. Mäkelä, 2017: Evaluation of thunderstorm predictors for finland using reanalyses and neural networks. *Journal of Applied Meteorology and Climatology*, **56 (8)**, 2335–2352.

Van Vleck, J., 1947a: The absorption of microwaves by oxygen. *Physical Review*, **71 (7)**, 413.

Van Vleck, J., 1947b: The absorption of microwaves by uncondensed water vapor. *Physical Review*, **71 (7)**, 425.

Wapler, K., 2017: The life-cycle of hailstorms: Lightning, radar reflectivity and rotation characteristics. *Atmospheric Research*, **193**, 60–72.

Weisman, M. L., and J. B. Klemp, 1982: The dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Monthly Weather Review*, **110 (6)**, 504–520.

Wilhelm, J., 2022: *Einfluss atmosphärischer Umgebungsbedingungen auf den Lebenszyklus konvektiver Zellen in der Echtzeit-Vorhersage*. KIT Scientific Publishing.

Wilhelm, J., S. Mohr, H. J. Punge, B. Mühr, M. Schmidberger, J. E. Daniell, K. M. Bedka, and M. Kunz, 2021: Severe thunderstorms with large hail across germany in june 2019. *Weather*, **76 (7)**, 228–237.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, Vol. 100. Academic press.

Willmott, C. J., and K. Matsuura, 2005: Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, **30 (1)**, 79–82.

Zhang.Y, and Niu.J, 2009: The characteristics of thunderstorm frequency variation and their possible relation with the adjustment of crop distribution in the leizhou peninsula. *JOURNAL OF TROPICAL METEOROLOGY*, **15 (1)**, 89–92.

Zrnić, D., R. Doviak, V. Melnikov, and I. Ivić, 2014: Signal design to suppress coupling in the polarimetric phased array radar. *Journal of Atmospheric and Oceanic Technology*, **31 (5)**, 1063–1077.

Acknowledgments

A year ago, when I embarked on this intellectual journey and took responsibility for my thesis, the end seemed a distant prospect, almost an illusion. My memories of those early days, filled with uncertainty and self-doubt, are still fresh. I remember questioning my abilities, wondering if I could do well to the research at hand. But now, at the end of a fulfilling year, my heart swells with satisfaction and joy. I am satisfied with the wealth of knowledge and insight I have accumulated. Research, once a looming task, has turned out to be a journey of self-discovery and personal growth. It has shown me the true nature of academic inquiry and the profound satisfaction that comes from it.

This growth of my academic and personal self, would not have been possible without the wise guidance and heartfelt advice of my esteemed research supervisors, Prof. Dr. Michael Kunz, Prof. Dr. Peter Knippertz, Dr. Julian Quinting and Dr. Jannik Wihelm. Their expertise, both academic and spiritual, has been a beacon illuminating my path during this transformative journey. For this, I am eternally grateful from the bottom of my heart.

I am deeply grateful to Michael for his professional feedbacks and advice and especially for his great patience. I still remember him sitting down face to face with me on more than one occasion and explaining in detail all the research questions that I was unfamiliar with, even though he was very busy. I am also thankful to him for giving me the opportunity to be part of his research team to broaden my research horizons. His rigorous and conscientious altitude to research issues was infectious in every conversation I had with him and made me respect him very much.

I also have a lot to appreciate Julian for. Thank you so much for always making time for me when I needed your support. And thank you for all the countless useful comments and ideas you gave me on machine learning, pulling me out of the vortex of overthinking to focus on the most critical aspects of problem solving. Thank you for telling me that even senior researchers also need to see the creativity of their students and look outside their bubble to see the something else. These words have encouraged me to experiment with my own ideas and to dare to face failure.

And Jannik, I am impressed by your knowledge of statistics. I also very appreciate that you give me a lot of academic advice without forgetting to ask me about my feelings. Thank you for your patience and for making me really enjoy those times when we discussed problems together until they were clarified. Your caring instructions often light my sights in the research direction.

Erklärung

Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet zu haben.

Karlsruhe, den 03.07.2023

Siyu Li