Identifying Lightning Processes in ERA5 Soundings with Deep Learning

Gregor Ehrensperger[1, 2], Thorsten Simon[2], Georg J. Mayr[2], and Tobias Hell[1]

[1]Data Lab Hell GmbH, Austria
[2]Department of Atmospheric and Cryospheric Sciences, University of Innsbruck

Correspondence: Gregor Ehrensperger (gregor.ehrensperger@uibk.ac.at)

Abstract. Atmospheric environments favorable for lightning and convection are commonly represented by proxies or parameterizations based on expert knowledge such as CAPE, wind ~~shears~~shear, charge separation, or combinations thereof. Recent developments in the field of ~~machine learning,~~ high resolution reanalyses, ~~and~~ accurate lightning observations, machine learning (ML) and explainable artificial intelligence (XAI) open possibilities for identifying tailored proxies without prior expert knowledge.

~~To identify vertical profiles favorable for lightning,~~ This study utilizes a deep neural network ~~links~~ trained to match temporally and vertically well-resolved ERA5 ~~vertical profiles~~ soundings of cloud physics, ~~mass field variables and wind to lightning location data from the~~ mass-field, and wind-field variables with lightning observations from the *Austrian Lightning ~~Detection~~ & Information & Detection System* (ALDIS)~~, which has been transformed to a binary target variable labeling~~. The ML model only receives the raw model atmosphere data as inputs, without incorporating any expert parameters or proxies derived from the model levels. Using and adapting appropriate XAI methods, it is then demonstrated how the inner workings of this well-performing deep learning model can be uncovered to identify physically meaningful patterns within the ERA5 ~~cells as *cells with lightning activity* and *cells without lightning activity*.~~ soundings that describe lightning processes.

The ERA5 parameters are taken on model levels beyond the tropopause forming an input layer of approx. 670 features~~. The data of 2010–2018 serve as training/validation.~~, the lightning data are transformed to a binary target variable labeling the spatio-temporal ERA5 grid cells as *cells with lightning activity* and *cells without lightning activity*.

~~On independent test data, 2019, the deep network outperforms a reference with features based on meteorological expertise. SHAP values~~ Scaled SHAP values are introduced to highlight the atmospheric processes learned by the ~~network which~~ neural network and show that the model identifies cloud ice and snow content in the upper and mid-troposphere as very relevant features. As these patterns correspond to the separation of charge in thunderstorm ~~cloud~~clouds, the deep learning model can serve as a physically meaningful description of lightning.

~~Depending on the region, the neural network also exploits the vertical wind or mass profiles~~ The scaled SHAP values also reveal that, depending on the location, the model additionally learns to correctly classify cells with lightning activity by exploiting mass-field or wind-field variables.

This approach also showcases how XAI can be used to accelerate knowledge discovery in areas where expert knowledge is still scarce.

1 Introduction

Lightning affects many fields of our everyday's life. Cloud-to-ground flashes might hit infrastructure such as wind turbines (Becerra et al., 2018) and power lines (Cummins et al., 1998) and thus cause power outages. Humans might get injured (Ritenour et al., 2008) or even die (Holle, 2016) after being hit by lightning. Wildfires (Reineking et al., 2010) release carbon dioxide into the climate system, and thus limit the biosphere's capacity to store carbon dioxide. Lightning also affects the climate system by producing nitrogen oxides which play a key role in ozone conversion and acid rain production (DeCaria et al., 2005). Ozone is an important greenhouse gas and changes in concentration can lead to warming or cooling of the atmosphere. Thus, understanding of lightning is also an important factor in climate change research (Finney et al., 2018).

Given lightning's impact and that an average of 46 flashes are occurring around the globe every second (Cecil et al., 2014) it is desirable to have models of the atmosphere capable to simulate lightning and its underlying dynamic processes down to the resolved scales of the numeric model. Beyond the resolved scales one relies on so called proxies *or* ~~parameterization~~ parameterizations to further describe lightning. The term *proxy* is commonly used for quantities derived from atmospheric model output *after* ~~numeric computations~~the simulation has run. *Parameterizations* ~~mean the description of~~ diagnose lightning *while* ~~numeric computations of the atmosphere model~~ the model is running and hence can feed back on the simulation.

Proxies are frequently applied to assess historic and future behavior of convection and lightning. Popular proxies are cloud top height (Price and Rind, 1992), cloud ice flux (Finney et al., 2014), CAPE times precipitation (Romps et al., 2018), or the lightning potential index (Brisson et al., 2021). Though, these proxies perform reasonably ~~good~~ well (Tippett et al., 2019), there is a need for more complex or holistic proxies, as the behavior of lightning in a changing climate is still uncertain (Murray, 2018). Another application ~~that makes clear that more research on the description of lightning is needed in the field of~~ highlighting the need for further research on lightning description is operational weather forecasting. Experience ~~shows~~indicates, for instance, that CAPE needs to be adapted to local conditions in order to perform well ~~(?)~~ (Groenemeijer et al., 2019).

Parameterizations are an internal part of numeric models, as they emulate sub-scale processes that cannot be resolved due the discretization of governing equations. Therefore, the emulated processes give feedback to the other processes, also on larger scales, within the atmospheric model. For instance, Tost et al. (2007) showed that modeled nitrogen oxide is sensitive to lightning parameterizations in numerical models. Next to the classic description of lightning using cloud top height (Price and Rind, 1992), parameterizations have been developed using polynomial regression (Allen and Pickering, 2002) and schemes based on hydrometeors in the mixed-phase region which is important for cloud-resolving models (McCaul et al., 2009). A comparison of several parameterizations using a superparameterized model is given by Charn and Parishani (2021). Recently, the ECMWF launched a product for total lightning densities expressed as a function of hydrometeors contents, CAPE, and (convective) cloud-base height output by the convective parameterization (Lopez, 2016).

In recent years~~machine learning (ML)~~ , also machine learning approaches have been proposed to describe convection and lightning. Forty preselected ~~single level~~ single-level parameters from ERA5 were processed by artificial neural networks and gradient boosting machines ~~for~~ to study lightning in parts of Europe and Sri Lanka (Ukkonen et al., 2017; Ukkonen and Mäkelä,

2019). ~~The authors also bring up the idea to feed ERA5's model level parameters directly to an appropriate ML tool, i.e. neural network. Other studies tested~~ Other studies evaluated random forests for regions such as the Hubei Province in China (Shi et al., 2022) or the Southern Great Plains (Shan et al., 2023) and generalized additive models (GAM) for the European Alps (Simon et al., 2023). All these studies confirm that the use of ML approaches for the description of lightning is promising. ~~In concurrent research, also~~

Very recently, *explainable artificial intelligence* (XAI) techniques are used to move towards understanding the underlying reasoning of complex AI models and ~~shows~~ show encouraging results in various Earth System Sciences applications (Barnes et al., 2020; Dutta and Pal, 2022; Hilburn et al., 2021; Mayer and Barnes, 2021; Stirnberg et al., 2021; Toms et al., 2021). Specifically, Silva et al. (2022) use XGBoost classification trees to explore when the NASA Goddard Earth Observing System model of lightning flash occurrence shows weaknesses and apply *Shapley additive explanations* (SHAP) to describe which meteorological drivers are related to the model errors. They found that these errors are strongly related to convection in the atmosphere and certain characteristics of the land surface.

This paper builds upon these studies and ~~aims at finding a *holistic* description of lightning. Supervised deep learning harvests temporally and vertically well resolved ERA5 soundings of atmospheric dynamics and cloud physics to explain observations from the *Austrian Lightning Information & Detection System* (ALDIS). The pattern found in ERA5 serve as proxy, but could also guide towards a parameterization of lightning. Using~~ demonstrates the use of explainable artificial intelligence to discover potential proxies favorable for lighting directly from raw model level atmospheric data. Unlike prior research (Ukkonen et al., 2017; Ukkonen and Mäkelä, 2019; Shi et al., 2022; Shan et al., 2023; Simon et al., 2023) that applied machine learning to classify lightning occurrence using preselected proxies derived from atmospheric parameters by experts, this work directly exploits the raw ERA5 ~~on model levels comes with the benefit that a complete picture of the atmosphere is considered to find patterns explaining lightning. As the approach sees the *raw model atmosphere*, no expert parameters diagnosed from the~~ model level data and targets at finding such proxies. Using model ~~levels are used as inputs, the study also answers whether deep learning can identify physically meaningful patterns within the ERA5 sounding to describe lightning processes.~~ level data directly offers two key benefits. First, it reduces the risk of overlooking potentially significant atmospheric conditions that could be missed when concentrating solely on preselected proxies. Second, it provides a comprehensive view of the vertical atmospheric layers, requiring less meteorological expertise to prepare the input data. This approach, however, increases the dimensionality of the input layer with highly correlated features along the vertical axis, making commonly used feature importance graphs hard to interpret. Inspired by the use of SHAP values in imaging tasks, this work employs SHAP values to reason on model levels directly. Due to the high dimensionality of the input, out of the box plotting routines are not feasible for interpreting SHAP values in this context. Therefore, the obtained SHAP values are aggregated to provide a more global understanding of a feature's contribution to the final model output. To improve explainability, *scaled SHAP values* are introduced to align the SHAP values across all grid cells. The median, as well as the 25th and 75th percentiles of these scaled SHAP values are then visualized along the vertical profiles, aiding the interpretation of the patterns exploited by the model.

~~The region of interest are the eastern Alps which are characterized by complex terrain. Atmospheric dynamics on a gamut of scales interacting with topography, which lead to various meso-scale (Feldmann et al., 2021) and local processes (Houze, 2012)~~

~~that can trigger convection and lightning.~~ This study focuses on lightning during the peak phase of the warm season (June, July, August) which differs fundamentally in the underlying dynamic processes to lightning during the cold season (Morgenstern et al., 2022). ~~Morgenstern et al. (2023) show that there are different environments, either dominated by wind-field or mass-field variables, that favor lightningdepending on the region~~

The region of interest are the Eastern Alps which are characterized by complex terrain. Atmospheric dynamics on a gamut of scales interact with topography, leading to various meso-scale processes (Feldmann et al., 2021) and local processes (Houze, 2012) that can trigger convection and lightning.

~~The~~ This paper is structured as follows. ~~Section 2~~ Section 2 presents both the lightning detection data and the atmospheric reanalyses~~, which both enter the supervised deep learning approach and a reference model (Sect. 3). Additionally, Section 3 illustrates the methods to analyze performance and explainability of~~ . Section 3 describes the two modelling approaches and elaborates on the XAI method used to interpret the patterns identified by the deep learning ~~approach~~model. The results of these analyses are given in ~~Section 4. Section 5~~ Section 4. Section 5 discusses the physical patterns identified by the ~~method, scrutinizes further applications and research that is made possible with the novel insights and~~ methods, highlights future applications and finally concludes the study.

2 Data

Two data sets build the foundation for this supervised machine learning task. First, the observational data from the lightning location system ALDIS (~~Sect. 2.1~~Section 2.1) is used to derive the labels distinguishing cells with and without lightning activity. Second, pseudo soundings from ERA5 (~~Sect. 2.2~~Section 2.2) serve as input for the deep learning approach. Spatially, the grid centers range from $8.25°E$ to $16.75°E$ and from $45.25°N$ to $49.75°N$.

Temporally, data for the meteorological summers (June, July, August) from 2010 to 2019 are available. The data of 2010–2018 serve as training/validation[1], the data from 2019 is reserved as truly independent test data.

2.1 Lightning Detection Data

The Austrian Lightning Detection & Information System (ALDIS) is part of the European ~~effort EUCLID~~Cooperation for Lightning Detection (EUCLID) (Schulz et al., 2016). Cloud-to-ground flashes with a current ~~of $>$ 15 kA or $<$ −2 kA~~ greater than 15 kA or smaller than −2 kA are aggregated to the spatio-temporal grid cells of ERA5 (~~Sect. 2.2~~Section 2.2). Each cell has a horizontal extent of ~~approx. 30 km × 30 km~~ $0.25° \times 0.25°$ and temporally of one hour. If at least one flash has been detected in such a grid cell, then the cell is labeled as *cell with lightning activity*. Otherwise, if not a single flash has been detected, the cell is labeled as *cell without lightning activity*.

[1] Data is split based on distinct days. 20% of these distinct days are used for validation, while the remaining 80% serve as training dataset.

Table 1. ERA5 parameters on model levels.

Name	Short Name	Units	Parameter ID
Temperature	t	~~K~~ K	130
Specific humidity	q	~~kg kg^{-1}~~ kg kg^{-1}	133
U component of wind	u	~~m s^{-1}~~ m s^{-1}	131
V component of wind	v	~~m s^{-1}~~ m s^{-1}	132
Vertical velocity	w	~~Pa s^{-1}~~ Pa s^{-1}	135
Specific rain water content	crwc	~~kg kg^{-1}~~ kg kg^{-1}	75
Specific snow water content	cswc	~~kg kg^{-1}~~ kg kg^{-1}	76
Specific cloud liquid water content	clwc	~~kg kg^{-1}~~ kg kg^{-1}	246
Specific cloud ice water content	ciwc	~~kg kg^{-1}~~ kg kg^{-1}	247

2.2 Atmospheric Reanalysis

ECMWF's fifth reanalyses, ERA5 (Hersbach et al., 2020), is available at a horizontal resolution of ~~0.25°~~ $0.25° \times 0.25°$ (in the region of interest this corresponds to approx. $19\,\mathrm{km} \times 28\,\mathrm{km}$) and temporally of ~~1~~ one hour. Vertically it consists of 137 hybrid model levels that align with topography near ground and approach isobars in the upper atmosphere~~(see)~~[2]. On these model levels nine parameters (~~Tab. 1~~Table 1) are available to describe the state of the atmosphere. ~~Next~~ In addition to classical parameters ~~,~~such as temperature, specific humidity and three-dimensional winds, ERA5 provides a description of liquid and solid water particles in clouds, i.e. the specific content of ice, snow (including graupel), liquid water, and rain. For this study, these parameters are used on the lowest 74 model levels, spanning from level ~~number~~ 64 (approx. $15\,000\,\mathrm{m}$ geopotential height) to level ~~number~~ 137 ($10\,\mathrm{m}$ above ground).

2.3 Composition of Datasets

The two data sets are merged in order to obtain a tabular data shape. Each row of this tabular data refers to a spatio-temporal grid cell. Thus, it can be indexed by the longitude and latitude of its center as well as its hourly time stamp. Each row is either labeled as cell with lightning activity or without lightning activity. The nine ERA5 parameters (~~Tab. 1~~Table 1) on their 74 model levels enter the tabular data such that each resulting column refers to an *individual* parameter on an *individual* level, making up a total of ~~$9 \times 74 = 666$~~ $9 \cdot 74 = 666$ ERA5 feature columns. Further, each row is complemented with the information of the *hour of the day* and *day of the season* to account for diurnal and seasonal variations, respectively. Finally, the model topography[3] is added as another column.

[2] See https://confluence.ecmwf.int/display/UDOC/L137+model+level+definitions.

[3] The topography is represented by a single scalar value: the geopotential height from model level 137, which is the layer adjacent to the Earth's surface at the specified grid point.

3 Methods

To avoid incorporating expert knowledge by using specialized deep learning architectures and to efficiently handle a large number of input features, a classical fully connected ~~deep~~ neural network (~~Sect. 3.1~~Section 3.1) is used~~to fit a model which capable of distinguishing whether a specific spatio-temporal grid cell corresponds to a lightning cell~~. To make sure that the neural network can model lightning sufficiently well and is worth being analyzed, the resulting outputs are compared to those of a state-of-the-art reference model (~~Sect. 3.2~~Section 3.2) on unseen test data. Finally, insights into the patterns exploited by the trained model are gained by applying Shapley additive explanations (~~Sect. 3.3~~Section 3.3).

3.1 Deep Learning Approach

A ~~rather general~~ fully connected neural network was designed, consisting of eight hidden layers with $512 \times 512 \times 512 \times 512 \times 128 \times 128 \times 128 \times 16$ nodes. Leaky rectified linear unit (leaky ReLU) is used as activation function for all hidden layers. The input dimension is predetermined by the number of input features and thus equals 671 (nine atmospheric variables on 74 levels, longitude, latitude, hour of the day, day of the season, and topography). The dimension of the output layer equals one, as it solely classifies whether the cell is with or without lightning activity. The model output is activated with the sigmoid function. ~~The input features are scaled, such that the nine atmospheric~~

Prior to training, the input variables are standardized~~by considering the~~. For each of the atmospheric variables $v \in \{\text{ciwc}, \text{clwc}, \text{crwc}, \text{cswc}$ the mean μ_v and standard deviation σ_v are calculated over all 74 ~~levels altogether, prior training.~~ model levels together, but separately for each of the nine variables.

To prevent the model from overfitting, dropout (Srivastava et al., 2014) with a value of 0.15 and early stopping with a patience of ten epochs are applied. Binary cross entropy serves as loss function with a weight of approximately 41 for positive events (flash occurrences) to address for the highly imbalanced data set.

3.2 Reference Model

For reference a generalized additive model ~~(GAM, Wood, 2017) is used~~ (GAM) (Wood, 2017) is used and fitted using an algorithm tailored for gigadata (Wood et al., 2017). This model is trained on longitude, latitude, hour of the day, day of the season, topography and the atmospheric variables listed in ~~Tab. 2~~Table 2, which were derived from ERA5 soundings on meteorological expertise (Simon et al., 2023).

Thus, the input dimension for the reference model is only 15. ~~The GAM is fitted using an algorithm tailored for gigadata (Wood et al., 2017).~~

3.3 Explainability

While generalized additive models are interpretable by users (Lou et al., 2012), interpretability research of deep neural networks still suffers many gaps (Zhang et al., 2021). ~~Deep~~ In this work SHAP (Lundberg and Lee, 2017) is utilized to gain insights

Table 2. The reference model is trained using the following ten atmospheric variables.

Description	Short Name
Convective available potential energy	`cape`
Binary indicator whether cloud is present	`cloud_exists`
Convective precipitation	`cp`
Mass of specific snow water content between the $-20°C$ and $-40°C$ isotherms	`cswc2040`
Cloud top height in height above ground	`cth`
Instantaneous surface sensible heat flux	`ishf`
Medium cloud cover	`mcc`
Total column supercooled liquid	`tcslw`
Mass of water vapor between the $-10°C$ and $-20°C$ isotherms	`wvc1020`
Two meter temperature	`2t`

into the patterns exploited by the neural network from ~~Section 5~~ Section 3.1 and to understand the features contributing to the classification of a spatio-temporal cell as one exhibiting lightning activity.

SHAP is a game theoretic approach which can be used to explain the relation of input and output of any machine learning model. It ~~uses~~ follows the concept of Shapley values (Shapley, 1952) to provide local interpretability by computing feature attributions ~~which~~ that lead to the model's output for a given input. Unfortunately, the computation time for calculating ~~the~~ exact Shapley values grows exponentially with the number of input features~~. Common implementations for computing Shapley values use simplifications to ensure computational feasibility~~, leading to various ways in which Shapley values are operationalized (Sundararajan and Najmi, 2020; Chen et al., 2023). The two main approaches, observational and interventional, differ in the way they sample dropped input features to attribute for the difference between the model output and the expectation caused by the removed feature (Chen et al., 2020). While there is an ongoing debate about which approach is preferable (Chen et al., 2020), Janzing et al. (2020) argues, supported by experiments, that the observational approach is flawed and interventional provides the correct notion of dropping features.

This work applies Deep SHAP[4] (Lundberg and Lee, 2017) which is a model agnostic method that leverages extra knowledge about the nature of deep neural networks to approximate Shapley values more efficiently~~but also assumes independence among the input features. As in various applications, this property obviously is not fulfilled in the given data set but using a more accurate approximation as described in Aas et al. (2021) is not feasible with the large number of input features . Despite this formal prerequisite, SHAP values are successfully utilized in a variety of imaging tasks (e.g. in the medical field (van der Velden et al., 2022)) although the independence assumption is also heavily violated when images are used as input and every pixel serves as an individual input feature. Understanding the raw ERA5 vertical profiles as a collection of 1D-images it is safe to assume that Deep SHAP leads to sufficiently good approximations to the precise Shapley values and can therefore be applied to gain insights into the neural network trained in this study~~. The input features in this work are highly

[4] Provided by the `DeepExplainer` class within the Python package `shap`.

correlated, particularly along the vertical profiles within a single variable. Deep SHAP belongs to the family of interventional methods, thus effectively identifies the features that the model genuinely uses to generate a specific output, even in the presence of correlated inputs.

4 Results

This section first ~~investigates~~ evaluates the performance of the deep learning approach ~~by comparing its output on unseen test data against observations and the output of~~ and compares it to the reference model (~~Sect. 4.1~~Section 4.1). Next, the application of SHAP ~~allows to gain insights about~~ provides insights into the vertical profiles ~~exploited by~~ that the neural network ~~which indicate the occurrence of lightning (Sect. 4.2~~found to be favorable for lightning (Section 4.2).

4.1 Performance of the deep learning approach

The neural network is trained as described in ~~Sect. 3.1~~ Section 3.1 to distinguish whether a given spatio-temporal cell is a cell with or without lightning activity. To map the ~~output of the model~~model's output to a binary category, a threshold has to be defined. ~~The~~ Due to the highly imbalanced nature of the given data set, this threshold is determined by maximizing the F_1 score, which ~~symmetrically represents~~ balances precision and recall~~in one metric~~, on the validation set.

This study aims at finding the atmospheric patterns exploited by the neural network to classify cells being with or without lightning, making the strategy and exact choice of threshold less critical. However, before analyzing the inner workings of the model it is essential to ensure that the trained model's performance is comparable to or even better than a state-of-the-art reference model.

The reference model is fitted as described in ~~Sect. 3.2~~ Section 3.2 and the threshold is computed following the same procedure.

~~The resulting confusion matrices are displayed in Tab. 3. The~~ From the confusion matrices displayed in Table 3 it can be concluded that the neural network slightly outperforms the reference model in every category of the confusion matrix on previously unseen test data (year 2019). This ~~can also be seen~~ is further supported by comparing the *Matthew correlation coefficients* (mcc) of the two models, where $+1$ represents a perfect match between model output and observations, and 0 indicates no better than random guessing. The deep learning model has an mcc of approximately 0.278~~and the reference model 0.237.~~

~~Previous studies have shown that descriptions of lightning based on ML models reproduce the observed diurnal cycle more realistically than simple proxies such as CAPE (Fig 2. in Simon et al., 2023). To investigate the ability of the introduced model to reproduce the diurnal cycle of lightning, the mean of the binary model output and the observations of the test year 2019 for each hour of the day is calculated for four different small subdomains (Fig.**??**). In this case, the model's threshold is calibrated to align the average predicted and observed lightning frequencies of the validation set.~~

~~The comparison reveals a good match of the shapes of the modeled and observed diurnal cycles. In particular, the transition from the low values in the morning to the peak in the afternoons are well reproduced. For three of the subdomains (Flatlands,~~

Table 3. Confusion matrices of the neural network model (left) and the reference model (right) on test year 2019.

		observed	
		yes	**no**
modeled	**yes**	14 ~~370~~ 372	61 ~~446~~ 431
	no	15 ~~768~~ 766	1 374 ~~741~~ 756

	observed	
	yes	**no**
yes	12 654	65 176
no	17 484	1 371 011

~~Northern Alpine Rim and Southern Alpine Rim), the model slightly overestimates the observed occurrence probabilities, which is a result of different mean occurrences of lightning in the validation (used for finding the threshold) and test data (plotted). The only larger deviation can be found during the late afternoon in the High Alps, where the model overestimates the observed diurnal cycle. It should be noted that the curves of~~ , while the reference model ~~are not directly comparable to those presented in (Fig. 2 in Simon et al., 2023), despite using the same model architecture and input variables. This discrepancy arises from the fundamentally different data strategies employed. Simon et al. (2023) implemented a cross-validation method across the entire dataset spanning 2010 to 2019, averaged the calibrated probabilities for lightning occurrences and evaluated the mean of the modeled diurnal cycles over the ten years of available data. In contrast, the current study exclusively utilizes data from the years 2010 to 2018 to fit the model and threshold. The model's performance and resulting diurnal cycles are then assessed using data exclusively from the full year 2019, which has previously not been seen~~has an mcc of 0.237.

~~Diurnal cycles of the probability of a lightning occurring in a spatio-temporal cell. The binary model outputs of the neural network and reference model, as well as the binary observations have been averaged over four subdomains.~~

4.2 Identifying patterns exploited by the deep learning model

The ~~good~~ performance of the deep learning approach ~~motivates a closer look at what~~ encourages a closer examination of the patterns the model has learned ~~in order to distinguish~~ to differentiate between cells with and without lightning activity. A sample is classified as having lightning activity when the model output exceeds the threshold ϕ.

SHAP values (~~Sect. 3.3~~Section 3.3) indicate which inputs the neural network is particularly interested in. ~~Since the goal is to find patterns which are valid throughout the full region used for training, and unbiased by the frequency of lightning within a specific spatial cell, the SHAP values are computed separately for each spatial cell.~~[5] Given a specific input, the SHAP values of all input features always sum up, with only ~~small~~ minor approximation errors, to the difference between ~~the~~ a base value (derived from the expected model output based on so-called background data) and the actual model output. To identify patterns that are consistent across the entire training region and not influenced by the frequency of lightning in specific spatial cells, SHAP values and corresponding background data are calculated and sampled separately for each spatial cell. Specifically, for each spatial cell, the background data consists of the complete set of samples without lightning activity from that cell. To better understand the underlying patterns, the SHAP values are then scaled by dividing them by the difference between the base value

[5] ~~In particular, within SHAP's `DeepExplainer` the full number of samples without lightning activity of the corresponding spatial cell are used as background data.~~

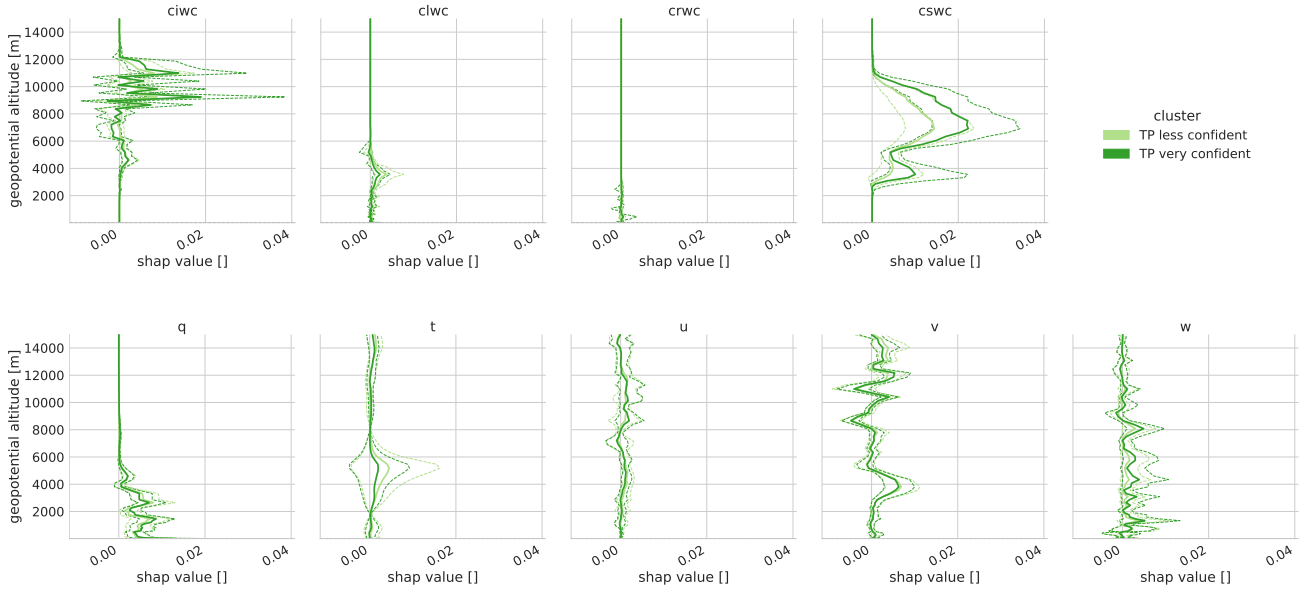

Figure 1. Scaled SHAP values for several variables (names on top of each subfigure) on correctly modeled lightning events (true positives). The two colors represent the confidence (stratified by median) of the network in its output. The dark green color summarizes the events where the network is very confident that a lightning event occurred. The light green color summarizes the events where the network still modeled correctly, but with less confidence. The solid lines show the median of all observations and the ~~colored areas~~ dashed lines highlight the ~~50% quantiles~~interquartile range.

of the corresponding spatial cell and the threshold (ϕ) at which a cell is classified as having lightning activity. This implies that the model classifies a ~~cell as having~~ sample to have lightning activity as soon as the scaled SHAP values sum up to one or more, regardless of the underlying base value ~~. All plots in this paper illustrate these scaled SHAP values.~~ and location.

Expressiveness is further improved by splitting the class of true positives into *less confident* and *very confident*. True positives with a model output in the interval $[\phi, \frac{1+\phi}{2})$ are considered *less confident true positives* and true positives with a model output in $[\frac{1+\phi}{2}, 1]$ are termed *very confident true positives*.

The aggregated results of the scaled SHAP values of correctly classified cells with lightning activity are visualized in ~~Fig. 1~~Fig. 1.

On average cloud ice (ciwc) and snow water content (cswc) contribute the most to the ~~network~~model's output. Also note that ciwc with its lighter-weighted ice crystals is particularly interesting at a geopotential height of approx. 8000 to 12000 m and cswc with its solid precipitation at approx. 3000 to 10000 m.

Taking a closer look (~~Fig. 2~~Fig. 2) at the ciwc and cswc at these altitudes, it is noticeable that the model exhibits greater confidence when ciwc and cswc values are substantially elevated. Furthermore, there is a tendency for the model to produce false positives during periods of high ciwc and cswc, while false negatives are more prevalent when these values are low compared to correctly classified lightning events.

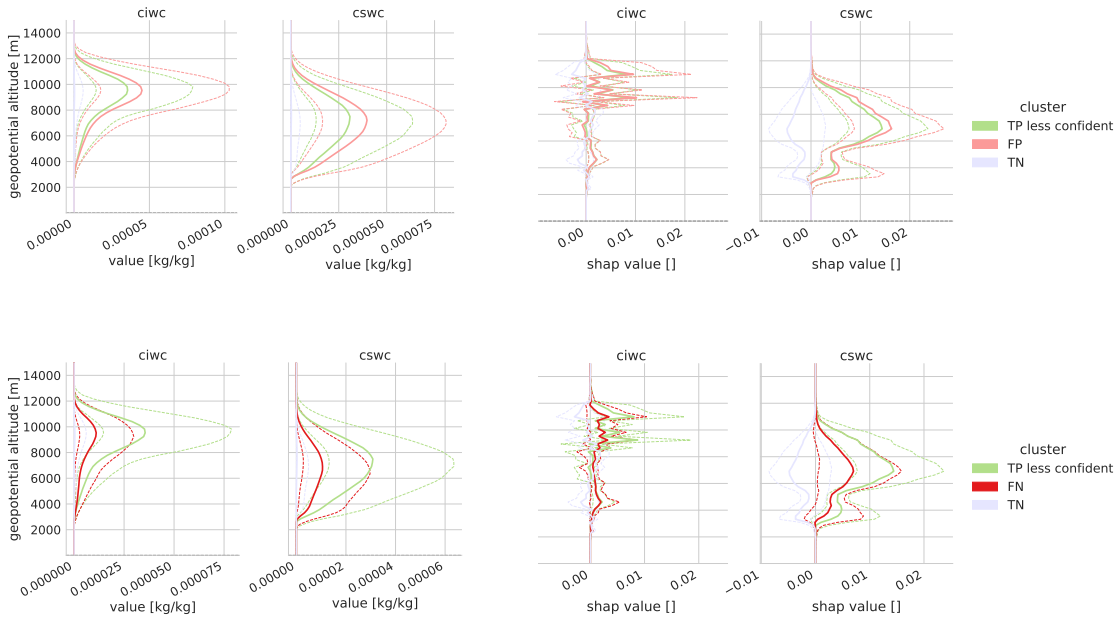

Figure 2. The two left columns display the vertical profiles of the real feature labels, while the two right columns present the vertical profiles of the scaled SHAP values. The upper row illustrates less confident true positives (TP) compared to false positives (FP), while the lower row illustrates less confident true positives compared to false negatives (FN). True negatives (TN) are also included for reference. The solid lines show the median of all observations and the dashed lines highlight the interquartile range.

While classifications where a cloudy atmosphere is the most dominantly exploited feature by the neural ~~networks~~ network are the majority, grouping the results into three categories, following Morgenstern et al. (2023), reveals additional patterns:

`cloud`: True positives where the sum of scaled SHAP values of ciwc, clwc, crwc and cswc over all model levels exceeds 0.5. Cloud-dominant cells with lightning activity are distributed across the entire region of interest, but are particularly abundant along the primary chain of the Alps.

`mass`: True positives where the sum of scaled SHAP values of q and t over all model levels exceeds 0.5. Mass-dominant cells are predominantly situated in Northern Italy and Slovenia.

`wind`: True positives where the sum of scaled SHAP values of u, v and w over all model levels exceeds 0.5. Wind-dominant cells are primarily concentrated in the northwestern region of the Italian flat terrain, the Po Plain.

Approximately 39.8% of the true positives belong to the cloud dominant, 2.6% to the mass dominant and 7.9% to the wind dominant class. Note that a single sample may belong to multiple groups or even none at all if the characteristics of cloud, mass, or wind are not distinctly pronounced.

Visualizing the vertical profiles of the ~~scaled SHAP values (Fig. ??) and the~~ real feature values (~~Fig. 3)of these three groups it becomes clear that~~ Fig. 3), their temperature profiles (t) are distinct. Events with high values for the mass-field ~~lightning is characterized by warmer temperatures in the troposphere, a less stable stratification and copious amounts~~ have warmer temperatures and their temperatures decrease more strongly with height than the other two classes. This indicates that less work is required to displace particles in the vertical thus making it more prone to produce thunderstorm clouds. Since the maximum possible amount of water vapor in the ~~lower troposphere. Larger amounts of latent heat released by condensation as indicated by large clwc values in the lower troposphere combined with a weaker stratification result in more CAPE that can be released~~ air before condensation occurs is exponentially related to temperature via the Clausius-Clapeyron equation, events with high values of the mass-field also have by far the largest values for specific humidity q, particularly in the part of the atmosphere closest to the surface. When that water vapor condenses as air is lifted from near the surface the latent heat released during this phase change will heat the air and thus decrease its density and make a further rise of the air parcels more likely. Since there is so much more water vapor available for a phase change than with the other two categories, one would expect the category with high mass field values to also have higher amounts of liquid and solid water (ciwc, ~~which carries solid particles (ciwc) higher into the troposphere. Heavier solid hydrometeors (cswc) peak further below.~~ clwc, crwc, cswc) at altitudes above the level where the phase change occurs. However, the opposite is the case. The explanation rests in the difference of the horizontal size of a grid cell of the ERA5 atmospheric reanalysis data, which is approximately $19\ \mathrm{km} \times 28\ \mathrm{km}$ in the region of interest, compared to the typical diameter of $5\ \mathrm{km}$ of the most frequent type of thunderstorms - single cell thunderstorms (Markowski and Richardson, 2010). ERA5 data are average values over the whole grid cell and when only one single-cell thunderstorm occurs in an ERA5 grid cell, the average cloud-variables will be low since most of the ERA5 grid cell is cloud-free. The lowest absolute values of vertical velocity of all three categories support this conclusion. The deep learning approach thus has learned lightning from single cell thunderstorms.

~~Ice crystals and solid hydrometeors in~~

The category with high wind-field ~~lightning, on the other hand, are not transported that far up into the troposphere and they both peak in a similar altitude range. The large-scale vertical velocity in the lower troposphere is high as is the horizontal wind speed~~ values has the coldest temperature (t) profiles of all three categories and – ~~particularly its southern component. Temperatures and consequently specific humidity q are lower and the stratification is stabler than for the mass-field lightning. All of this indicates forced lifting along (cold) frontsand topography~~because of the exponential relationship to maximum possible water vapor – also the lowest values of specific humidity (q) in the lower part of the atmosphere. Despite the least amount of water vapor available for condensation, this category has the largest amounts of cloud droplets (clwc) and of rain (crwc). Consequently such thunderstorms must occur in situations when most or all of an ERA5 grid cell is filled with clouds. Also, the absolute values of vertical velocity are largest of all three categories. The corresponding meteorological situations are large scale patterns of lifting in the atmosphere such as along cold fronts. Cold fronts in this region occur more frequently in the months between fall and spring, which explains why this category has the coldest temperatures. Also, cold fronts in this region typically occur in southwesterly flow downstream of the trough axis, which explains the exceptional large values of the ~~v-component~~ v-component of the wind. ~~Charge separation consequently~~ Since wind speed also increases most strongly with

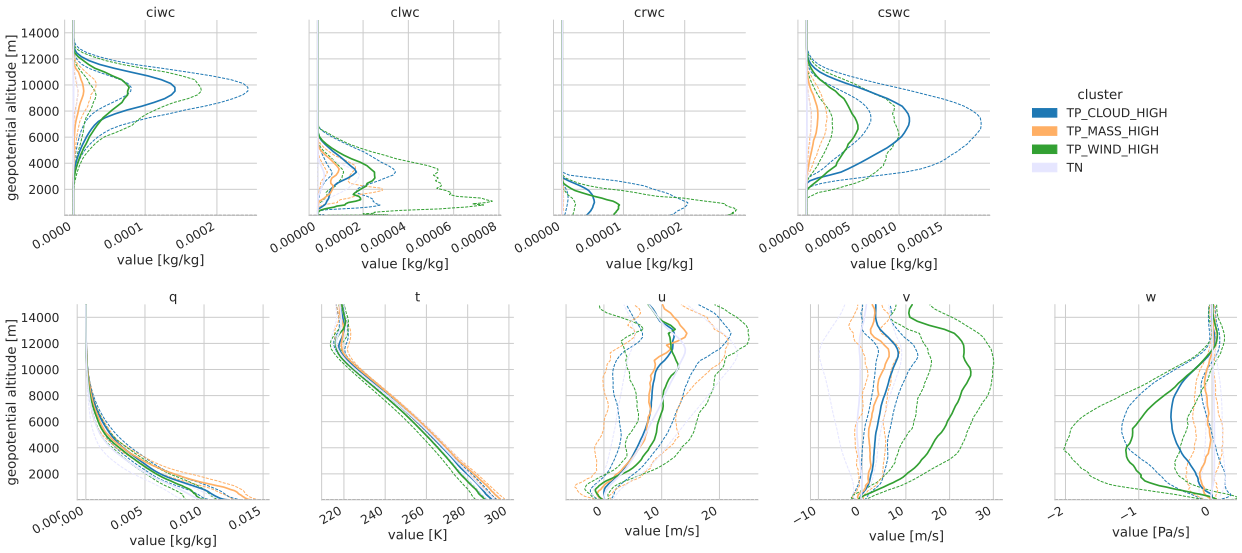

Figure 3. Vertical profiles of the real features per variable with colors indicating true negatives and different groups of true positives (cloud-, mass-, wind-dominant). The solid lines show the median of all observations and the dashed lines highlight the interquartile range. Note that in pressure coordinates, negative values of vertical velocity indicate upward motion.

height, charge separation occurs on a tilted instead of a nearly vertical path as in ~~mass field~~ mass-field lightning, having earned this type of lightning the name *tilted thunderstorm* (Brook et al., 1982; Takeuti et al., 1978; Takahashi et al., 2019; Wang et al., 2021).

~~Vertical profiles of the scaled SHAP-values (a) and real features (b) per variable with colors indicating true negatives and different groups of true positives (cloud-, mass-, wind-dominant).~~ The third category in Fig. 3 with high cloud-field variables has the largest amounts of solid water – ice crystals (ciwc), snow flakes and graupel (cswc) – but only the second largest amounts of liquid water (clwc, crwc). Also the vertical velocities are in between the other two categories. Therefore this category likely represents the meteorological situation of multicell and supercell thunderstorms (Markowski and Richardson, 2010), which have a larger footprint than single cell thunderstorms (the mass-field category) and will thus fill larger fractions of an ERA5 grid cell. This category could also contain cold front situations (the wind-field category) where the cold front occupies only parts of an ERA5 grid cell.

~~Following the approach of Morgenstern et al. (2023), Fig. 4 subdivides the cloud-dominant group into two subcategories:~~ To test the hypothesis that the category with high cloud-field values contains both of these situations, i.e. mass-field and wind-field dominated situations, we divide this category into a cloud-mass and a cloud-wind ~~. This~~ category in Fig. 4. This is an approach also taken by Morgenstern et al. (2023). The grouping is based on whether the aggregate of scaled SHAP values is greater for mass-related or wind-related parameters.

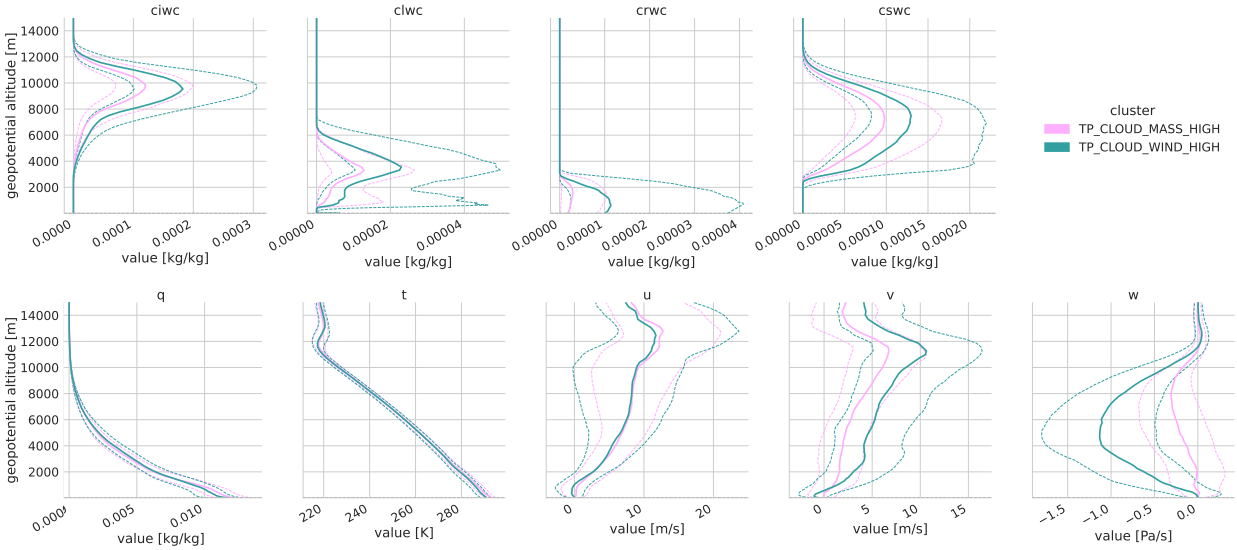

Figure 4. Vertical profiles of the ~~scaled SHAP-values (a) and~~ real features ~~(b)~~ per variable with colors indicating cloud-mass and cloud-wind dominant true positives. The solid lines show the median of all observations and the dashed lines highlight the interquartile range.

~~Lightning with large scaled SHAP and real values of the cloud variables seems to occur for both mass-field and wind-field lightning confirming the results from Morgenstern et al. (2022, 2023), who had used principal component analysis and clustering for identifying these categories.~~

~~Fig. ?? highlights the geographical regions, where cloud-mass-, cloud-wind-, mass-, or wind-dominant cells exhibiting lightning activity were classified. Cloud-dominant cells with lightning activity are distributed across the entire map, but are particularly abundant along the primary chain of the Alps. Mass-dominant cells are predominantly situated in Northern Italy and Slovenia. Wind-dominant cells are primarily concentrated in the northwestern region of the Italian flat terrain, the Po Plain~~And indeed, we find that the cloud-wind subcategory again has the largest amount of liquid water (clwc, crwc) and also larger values of the southerly wind component (v) indicative of the typical southwesterly flow for which (cold) fronts occur in this region. The cloud-wind category even has the higher solid water contents than the cloud-mass category indicating that even larger-sized thunderstorms in the absence of cold fronts do not always completely fill an ERA5 grid cell.

~~The count of true positive classifications stratified by the variable group that is most dominant for each geographical location. The data for the displayed topography layer is taken from TanDEM-X (Rizzoli et al., 2017).~~

4.3 Sample case study

~~A sample case on unseen test data illustrates how the model from the deep learning approach *sees* a specific weather event. In the afternoon of~~ Thunderstorms and lightning commonly exhibit linear organization along meteorological boundaries such as fronts or convergence zones. Our deep learning model, trained exclusively on individual vertical atmospheric profiles,

successfully identifies these linear structures without explicit knowledge of horizontal connections. A case study from June 20, 2019, ~~a weak upper level trough embedded in southwesterly flow passed over the Alps, whereas below crest height the flow was predominantly around the Alps. Lightning in the target area (Fig. 5) occured in the warm sector in a zone with the highest values of~~ demonstrates this capability. Two weak frontal systems occur in the region shown in Fig. 5. They are embedded within a region of high equivalent potential temperature ~~. Its accompanying front had just arrived on the west coast of Europe. The lightning model correctly identified lightning~~ (not shown). The bow-shaped front in the eastern half of the figure is more pronounced and extends over a larger part of the figure. The second one over Switzerland is only visible in the westernmost part of the figure. The deep learning approach model accurately reproduced the linear lightning pattern in the eastern ~~half of Fig. 5 while misclassifying several occurences in its western half~~region. However, it overestimated the width of the lightning zone and failed to capture its northernmost extent, as indicated by false positives (small green circles, Fig. 5). Nevertheless, the model exhibits deficiencies in reproducing the southwestern portion of the thunderstorm line over Switzerland, generating an erroneous linear feature further northward.

It is noteworthy that the threshold in this study was not chosen to perfectly calibrate the model, but instead to balance between precision and recall. Due to the heavy class imbalance, this generally results in overestimation.

5 Discussion and Conclusions

In this study~~a~~, the region of interest are the Eastern Alps, a region that offers a variety of atmospheric processes due to its complex terrain and is well understood (Simon et al., 2023; Morgenstern et al., 2023). This is important, because it allows for critical evaluation of the patterns uncovered by explainable AI methods and provides insights into whether this approach is suitable for accelerating scientific discovery in regions where knowledge is still scarce.

A neural network is trained on the vertical columns of raw ERA5 data without inducing any further expert knowledge about atmospheric processes to classify whether there was a lightning event or not. Then scaled SHAP values are used to explain which variables and vertical levels attribute the most to correct classifications of cells with lightning activity. As indicated in ~~Sect. 4.2~~Section 4.2, the specific snow water and ice water content significantly capture attention, with peak interest occurring at a geopotential height of approximately $4000\ \mathrm{m}$ and $7000\ \mathrm{m}$ (cswc), and at heights of $9000\ \mathrm{m}$ and $11000\ \mathrm{m}$ (ciwc) respectively. ~~The~~ Thus, the neural network discovered by itself the essential ingredient for lightning, namely charge separation. It occurs when ice ~~cyrstals~~ crystals (ciwc) and larger frozen particles (graupel, cswc) are present in the convective updraft. Once the graupel is sufficiently heavy, its velocity is smaller than the velocity of the rising ice crystals, and the collisions between ice crystals and graupel result in oppositely charged particles (Reynolds et al., 1957; Saunders et al., 2006). ~~Fig. 1 in Lopez (2016)~~ Lopez (2016, Fig. 1) shows the typical distribution of charges in a mature thunderstorm cloud. ~~However~~Additionally, it is noteworthy that the model seems to be particularly interested in the cloud ice water content at a height of 9000 ~~and~~ to $11000\ \mathrm{m}$ while recent literature usually ~~looks at~~ examines the cloud ice water content at $440\ \mathrm{hPa}$ (typically about $6000\ \mathrm{m}$) (Finney et al., 2014, 2018; Silva et al., 2022). Focusing on the region ~~close to the tropopause~~ between 9000 and $11000\ \mathrm{m}$ means that it is crucial to vent ice particles all the way up to the tropopause and form anvils, as is typical of ~~the~~ thunderstorm clouds.

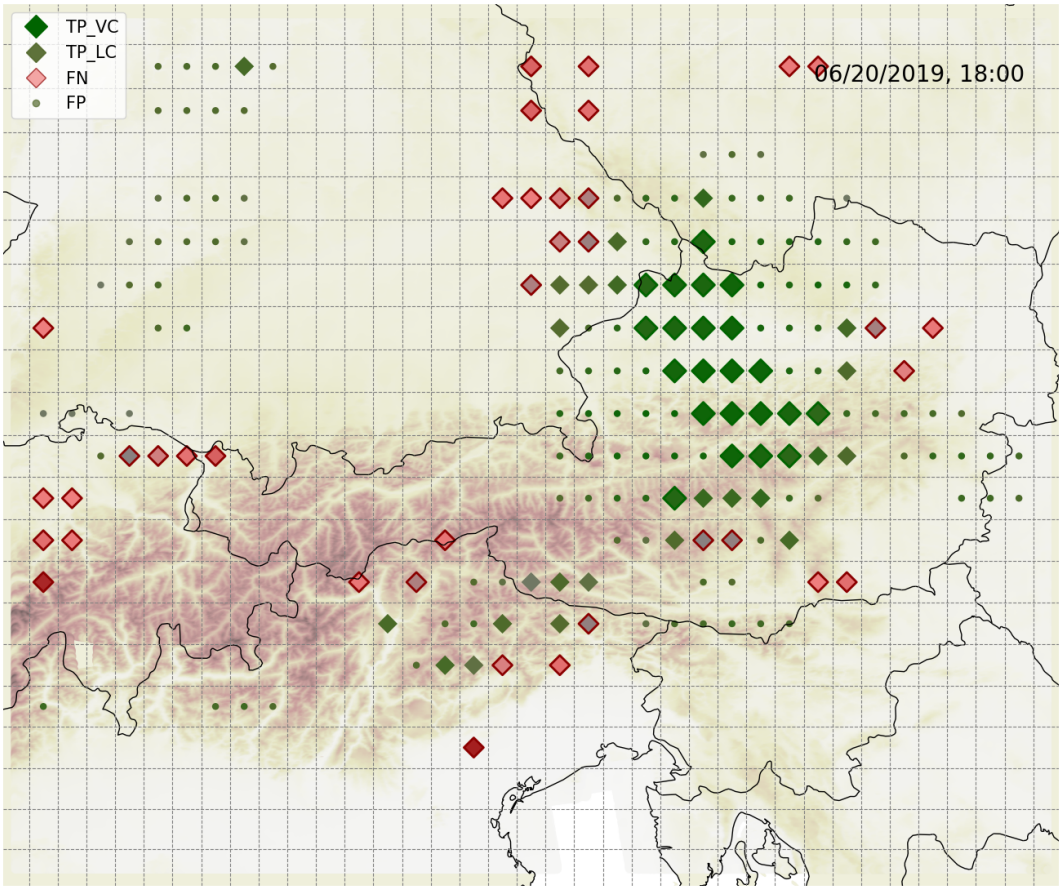

Figure 5. The map shows ERA5 grid cells with classifications of true positive (green diamonds), false negative (red diamonds) and false positive (dots) for the test data case June 20, 2019, in the hour before 18:00 UTC which is a case of the unseen test data. The size of the green diamonds indicates whether it is a *very* or *less* confident true positive. Low saturation of the red diamonds indicates that the output of the network was close to labeling the cell as one with lightning activity. The data for the displayed topography layer is taken from TanDEM-X (Rizzoli et al., 2017).

Moreover, the model leverages the presence of southerly winds and vertical updrafts as reliable indicators for lightning occurrence especially in the northwestern Po Plain. Additionally, high specific humidity below $4000\,\mathrm{m}$ serves as a robust proxy in the central and eastern Po Plain, as well as in the southern regions of the Slovenian Alps.

The case study in Section 4.3 demonstrates that, altough recall and precision of the neural network may appear to be low at first glance, the model effectively reproduces the general patterns of thunderstorms, despite overestimating and underestimating their extents. Similar observations were also made for many other examples not included in this manuscript.

The results in this work suggest promising future applications. Being able to train a neural network directly on atmospheric soundings with good ability to distinguish between cells with and without lightning activity, and then opening the black box may enable researchers to gain a better understanding of atmospheric processes in regions like e.g. equatorial Africa where ample studies are scarce (Chakraborty et al., 2022). The first ~~MGT-I~~ MTG-I satellite was launched on 13 December 2022 and will provide a lightning imager (Holmlund et al., 2021) which appears to be a promising source for the target variable. Furthermore, many existing models come with two very different parameterizations for ocean and land (Finney et al., 2014) and this inevitably leads to discontinuities in coastal areas. Also the reasons for the much lower lightning frequency over ocean are not as well understood yet. ~~Explainable AI~~ XAI might be a valuable building block in moving towards a more holistic understanding of the underlying atmospheric processes.

Using ML models to find parametrizations require them to be generalizable. In Ehrensperger et al. (2023) a similar model was trained on the same region but without using longitude, latitude and the day of the year as input features. While not giving the location to the model still provided a comparable performance, it enabled to evaluate the model on Continental Europe. The results show that the model is still able to perform comparably well on landcovered areas on previously unseen test data, demonstrating its ability to generalize across both time and location.

Future work might improve the results presented in this study. Here, a simple fully connected neural network is used and therefore the model loses information about the connectivity of the values along the levels of the vertical profiles. Using convolutional layers to process the profiles would, most likely, improve the results.

Convection and cloud processes are not purely vertical processes and thus ML parameterization greatly benefits from using multiple neighboring vertical atmospheric columns instead of a single column. Wang et al. (2022) work with $192\,\mathrm{km} \times 192\,\mathrm{km}$ grid cells to model, among others, subgrid zonal and meridional momentum flux due to vertical advection and suggest that a 3×3 subgrid could further improve the performance of the deep learning approach. ~~Here, a simple fully connected neural network is used and therefore the model loses information about the connectivity of the values along the levels of the vertical profiles. Using convolutional layers to process the profiles would, most likely, further improve the results. However, the goal of this work was to use a very simple machine learning approach to detect cells with lightning activity and then to disect the model to understand which atmospheric conditions the model has found to be typical for lightning. The input data was preprocessed with only very little meteorological expertise to ensure that the methodology is easily transferable to other regions of the earth where the understanding of lightning related atmospheric processes is still scarce.~~

Code and data availability. The software (version ~~1.1~~1.2; Python and R code) used to produce the results and plots in this manuscript is licenced under MIT and published on ~~GitHub https://github.com/noxthot/xai_lightningprocesses~~Zenodo (https://dx.doi.org/10.5281/zenodo.13907708)(Ehr
et al., 2024). The source code relies on two data sources:

1. ERA5 (Hersbach et al., 2020) data are available via the Climate Data Store (Hersbach et al., 2018, 2017). Scripts for sending the retrievals are included in the `data-preprocessing` directory of the ~~GitHub~~Zenodorepository (Ehrensperger et al., 2024).
2. ~~The~~ALDIS data (Schulz et al., 2016) ~~, which are the second important source of data, cannot be made available to the public. However, ALDIS data are available on request from ALDIS aldis@ove.at – fees may be charged.~~was aggregated to align with the spatio-temporal grid cells of ERA5 for use in this work. The transformed data is published in Simon et al. (2024).

Author contributions. **Gregor Ehrensperger**: Methodology, Software - model & explainable AI & plotting & data preparation, Writing – original draft. **Thorsten Simon**: Data curation, Software - reference model & plotting, Writing – original draft. **Georg Mayr**: Supervision, Writing - review & editing. **Tobias Hell**: Conceptualization, Methodology.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements. We are grateful for data support provided by Gerhard Diendorfer and Wolfgang Schulz from OVE-ALDIS. We also thank Deborah Morgenstern and Johannes Horak for their script to compute geopotential height on ERA5 model levels. ~~Also~~Additionally, we thank Johanna Rissbacher for contributing parts of ~~Fig. 5~~ Fig. 5 and the corresponding code. Furthermore, we greatly appreciate the insightful and constructive reviews from the anonymous reviewers, which have significantly enhanced the quality of this paper.

Financial support. This work was funded by the Austrian Science Fund (FWF, grant no. P 31836) and the Austrian Research Promotion Agency (FFG, grant no. 872656).

References

Aas, K., Jullum, M., and Løland, A.: Explaining Individual Predictions when Features are Dependent: More Accurate Approximations to Shapley Values, Artificial Intelligence, 298, 103 502, https://doi.org/10.1016/j.artint.2021.103502, 2021.

Allen, D. J. and Pickering, K. E.: Evaluation of Lightning Flash Rate Parameterizations for Use in a Global Chemical Transport Model, Journal of Geophysical Research: Atmospheres, 107, ACH 15–1–ACH 15–21, https://doi.org/10.1029/2002JD002066, 2002.

Barnes, E. A., Toms, B., Hurrell, J. W., Ebert-Uphoff, I., Anderson, C., and Anderson, D.: Indicator Patterns of Forced Change Learned by an Artificial Neural Network, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 195, https://doi.org/10.1029/2020MS002195, 2020.

Becerra, M., Long, M., Schulz, W., and Thottappillil, R.: On the Estimation of the Lightning Incidence to Offshore Wind Farms, Electric Power Systems Research, 157, 211–226, https://doi.org/10.1016/j.epsr.2017.12.008, 2018.

Brisson, E., Blahak, U., Lucas-Picher, P., Purr, C., and Ahrens, B.: Contrasting Lightning Projection Using the Lightning Potential Index Adapted in a Convection-Permitting Regional Climate Model, Climate Dynamics, 57, 2037–2051, https://doi.org/10.1007/s00382-021-05791-z, 2021.

Brook, M., Nakano, M., Krehbiel, P., and Takeuti, T.: The electrical structure of the hokuriku winter thunderstorms, Journal of Geophysical Research: Oceans, 87, 1207–1215, https://doi.org/10.1029/JC087iC02p01207, 1982.

Cecil, D. J., Buechler, D. E., and Blakeslee, R. J.: Gridded Lightning Climatology from TRMM-LIS and OTD: Dataset Description, Atmospheric Research, 135, 404–414, https://doi.org/10.1016/j.atmosres.2012.06.028, 2014.

Chakraborty, R., Menghal, P., Harshitha, M., and Sodunke, M.: Climatology of Lightning Activities Across the Equatorial African Region, in: 2022 3rd URSI Atlantic and Asia Pacific Radio Science Meeting (AT-AP-RASC), pp. 1–4, IEEE, https://doi.org/10.23919/AT-AP-RASC54737.2022.9814276, 2022.

Charn, A. B. and Parishani, H.: Predictive Proxies of Present and Future Lightning in a Superparameterized Model, Journal of Geophysical Research: Atmospheres, 126, https://doi.org/10.1029/2021JD035461, 2021.

Chen, H., Janizek, J. D., Lundberg, S., and Lee, S.-I.: True to the model or true to the data?, arXiv preprint arXiv:2006.16234, 2020.

Chen, H., Covert, I. C., Lundberg, S. M., and Lee, S.-I.: Algorithms to estimate Shapley value feature attributions, Nature Machine Intelligence, 5, 590–601, 2023.

Cummins, K., Krider, E., and Malone, M.: The US National Lightning Detection Network and Applications of Cloud-to-Ground Lightning Data by Electric Power Utilities, IEEE Transactions on Electromagnetic Compatibility, 40, 465–480, https://doi.org/10.1109/15.736207, 1998.

DeCaria, A. J., Pickering, K. E., Stenchikov, G. L., and Ott, L. E.: Lightning-Generated NO_x and its Impact on Tropospheric Ozone Production: A Three-Dimensional Modeling Study of a Stratosphere-Troposphere Experiment: Radiation, Aerosols and Ozone (STERAO-A) Thunderstorm, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2004JD005556, 2005.

Dutta, D. and Pal, S. K.: Interpretation of Black Box for Short-Term Predictions of Pre-Monsoon Cumulonimbus Cloud Events over Kolkata, Journal of Data, Information and Management, 4, 167–183, https://doi.org/10.1007/s42488-022-00071-9, 2022.

Ehrensperger, G., Hell, T., Mayr, G. J., and Simon, T.: Evaluating the generalization ability of a deep learning model trained to detect cloud-to-ground lightning on raw ERA5 data, Tech. rep., Copernicus Meetings, https://doi.org/10.5194/egusphere-egu23-15817, 2023.

Ehrensperger, G., Hell, T., Mayr, G., and Simon, T.: xai_lightningprocesses, https://doi.org/10.5281/zenodo.13907708, 2024.

Feldmann, M., Germann, U., Gabella, M., and Berne, A.: A Characterisation of Alpine Mesocyclone Occurrence, Weather and Climate Dynamics, 2, 1225–1244, https://doi.org/10.5194/wcd-2-1225-2021, 2021.

Finney, D. L., Doherty, R. M., Wild, O., Huntrieser, H., Pumphrey, H. C., and Blyth, A. M.: Using Cloud Ice Flux to Parametrise Large-Scale Lightning, Atmospheric Chemistry and Physics, 14, 12 665–12 682, https://doi.org/10.5194/acp-14-12665-2014, 2014.

Finney, D. L., Doherty, R. M., Wild, O., Stevenson, D. S., MacKenzie, I. A., and Blyth, A. M.: A Projected Decrease in Lightning under Climate Change, Nature Climate Change, 8, 210–213, https://doi.org/10.1038/s41558-018-0072-6, 2018.

Groenemeijer, P., Púcik, T., Tsonevsky, I., and Bechtold, P.: An Overview of Convective Available Potential Energy and Convective Inhibition provided by NWP models for operational forecasting, https://doi.org/10.21957/q392hofrl, 2019.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: Complete ERA5 from 1979: Fifth generation of ECMWF atmospheric reanalyses of the global climate, https://cds.climate.copernicus.eu/#!/home, accessed on 27-05-2021, 2017.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1959 to present, https://doi.org/10.24381/cds.adbb2d47, accessed on 16-02-2022, 2018.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 Global Reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hilburn, K. A., Ebert-Uphoff, I., and Miller, S. D.: Development and Interpretation of a Neural-Network-Based Synthetic Radar Reflectivity Estimator Using GOES-R Satellite Observations, Journal of Applied Meteorology and Climatology, 60, 3–21, https://doi.org/10.1175/JAMC-D-20-0084.1, 2021.

Holle, R. L.: A Summary of Recent National-Scale Lightning Fatality Studies, Weather, Climate, and Society, 8, 35–42, https://doi.org/10.1175/WCAS-D-15-0032.1, 2016.

Holmlund, K., Grandell, J., Schmetz, J., Stuhlmann, R., Bojkov, B., Munro, R., Lekouara, M., Coppens, D., Viticchie, B., August, T., Theodore, B., Watts, P., Dobber, M., Fowler, G., Bojinski, S., Schmid, A., Salonen, K., Tjemkes, S., Aminou, D., and Blythe, P.: Meteosat Third Generation (MTG): Continuation and Innovation of Observations from Geostationary Orbit, Bulletin of the American Meteorological Society, 102, 990–1015, https://doi.org/10.1175/BAMS-D-19-0304.1, 2021.

Houze, R. A.: Orographic Effects on Precipitating Clouds, Reviews of Geophysics, 50, 1–47, https://doi.org/10.1029/2011RG000365, 2012.

Janzing, D., Minorics, L., and Blöbaum, P.: Feature relevance quantification in explainable AI: A causal problem, in: International Conference on artificial intelligence and statistics, pp. 2907–2916, PMLR, 2020.

Lopez, P.: A Lightning Parameterization for the ECMWF Integrated Forecasting System, Monthly Weather Review, 144, 3057–3075, https://doi.org/10.1175/MWR-D-16-0026.1, 2016.

Lou, Y., Caruana, R., and Gehrke, J.: Intelligible Models for Classification and Regression, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, pp. 150–158, Association for Computing Machinery, New York, NY, USA, ISBN 9781450314626, https://doi.org/10.1145/2339530.2339556, 2012.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., vol. 30, Curran Associates, Inc., https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf, 2017.

Markowski, P. and Richardson, Y.: Mesoscale Meteorology in Midlatitudes, Wiley-Blackwell, ISBN 9780470742136, https://doi.org/10.1002/9780470682104, 2010.

Mayer, K. J. and Barnes, E. A.: Subseasonal Forecasts of Opportunity Identified by an Explainable Neural Network, Geophysical Research Letters, 48, e2020GL092 092, https://doi.org/10.1029/2020GL092092, 2021.

McCaul, E. W., Goodman, S. J., LaCasse, K. M., and Cecil, D. J.: Forecasting Lightning Threat Using Cloud-Resolving Model Simulations, Weather and Forecasting, 24, 709–729, https://doi.org/10.1175/2008WAF2222152.1, 2009.

Morgenstern, D., Stucke, I., Simon, T., Mayr, G. J., and Zeileis, A.: Differentiating Lightning in Winter and Summer with Characteristics of the Wind Field and Mass Field, Weather and Climate Dynamics, 3, 361–375, https://doi.org/10.5194/wcd-3-361-2022, 2022.

Morgenstern, D., Stucke, I., Mayr, G. J., Zeileis, A., and Simon, T.: Thunderstorm environments in Europe, Weather and Climate Dynamics, 4, 489–509, https://doi.org/10.5194/wcd-4-489-2023, 2023.

Murray, L. T.: An Uncertain Future for Lightning, Nature Climate Change, 8, 191–192, https://doi.org/10.1038/s41558-018-0094-0, 2018.

Price, C. and Rind, D.: A Simple Lightning Parameterization for Calculating Global Lightning Distributions, Journal of Geophysical Research: Atmospheres, 97, 9919–9933, https://doi.org/10.1029/92JD00719, 1992.

Reineking, B., Weibel, P., Conedera, M., and Bugmann, H.: Environmental Determinants of Lightning- v. Human-Induced Forest Fire Ignitions Differ in a Temperate Mountain Region of Switzerland, International Journal of Wildland Fire, 19, 541–557, https://doi.org/10.1071/WF08206, 2010.

Reynolds, S., Brook, M., and Gourley, M. F.: Thunderstorm Charge Separation, Journal of Atmospheric Sciences, 14, 426–436, https://doi.org/10.1175/1520-0469(1957)014<0426:TCS>2.0.CO;2, 1957.

Ritenour, A. E., Morton, M. J., McManus, J. G., Barillo, D. J., and Cancio, L. C.: Lightning Injury: A Review, Burns, 34, 585–594, https://doi.org/10.1016/j.burns.2007.11.006, 2008.

Rizzoli, P., Martone, M., Gonzalez, C., Wecklich, C., Borla Tridon, D., Bräutigam, B., Bachmann, M., Schulze, D., Fritz, T., Huber, M., Wessel, B., Krieger, G., Zink, M., and Moreira, A.: Generation and performance assessment of the global TanDEM-X digital elevation model, ISPRS Journal of Photogrammetry and Remote Sensing, 132, 119–139, https://doi.org/https://doi.org/10.1016/j.isprsjprs.2017.08.008, 2017.

Romps, D. M., Charn, A. B., Holzworth, R. H., Lawrence, W. E., Molinari, J., and Vollaro, D.: CAPE Times P Explains Lightning Over Land But Not the Land-Ocean Contrast, Geophysical Research Letters, 45, 12,623–12,630, https://doi.org/10.1029/2018GL080267, 2018.

Saunders, C. P. R., Bax-norman, H., Emersic, C., Avila, E. E., and Castellano, N. E.: Laboratory Studies of the Effect of Cloud Conditions on Graupel/Crystal Charge Transfer in Thunderstorm Electrification, Quarterly Journal of the Royal Meteorological Society, 132, 2653–2673, https://doi.org/10.1256/qj.05.218, 2006.

Schulz, W., Diendorfer, G., Pedeboy, S., and Poelman, D. R.: The European Lightning Location System EUCLID Part 1: Performance Analysis and Validation, Natural Hazards and Earth System Sciences, 16, 595–605, https://doi.org/10.5194/nhess-16-595-2016, 2016.

Shan, S., Allen, D., Li, Z., Pickering, K., and Lapierre, J.: Machine-learning-based investigation of the variables affecting summertime lightning occurrence over the Southern Great Plains, 23, 14 547–14 560, https://doi.org/10.5194/acp-23-14547-2023, 2023.

Shapley, L. S.: A Value for N-Person Games, RAND Corporation, Santa Monica, CA, https://doi.org/10.7249/P0295, 1952.

Shi, M., Zhang, W., Fan, P., Chen, Q., Liu, Z., Li, Q., and Liu, X.: Modelling Deep Convective Activity Using Lightning Clusters and Machine Learning, International Journal of Climatology, 42, 952–973, https://doi.org/10.1002/joc.7282, 2022.

Silva, S. J., Keller, C. A., and Hardin, J.: Using an Explainable Machine Learning Approach to Characterize Earth System Model Errors: Application of SHAP Analysis to Modeling Lightning Flash Occurrence, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 881, https://doi.org/10.1029/2021MS002881, 2022.

Simon, T., Mayr, G., Morgenstern, D., Umlauf, N., and Zeileis, A.: Amplification of annual and diurnal cycles of alpine lightning, Climate Dynamics, 61, 1–13, https://doi.org/10.1007/s00382-023-06786-8, 2023.

Simon, T., Schulz, W., Ehrensperger, G., and Mayr, G.: ALDIS cloud to ground lightning strike occurrence aggregated to spatiotemporal ERA5 cells (summer months 2010 to 2019), https://doi.org/10.5281/zenodo.13164463, accessed on 06-08-2024, 2024.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research, 15, 1929–1958, http://jmlr.org/papers/v15/srivastava14a.html, 2014.

Stirnberg, R., Cermak, J., Kotthaus, S., Haeffelin, M., Andersen, H., Fuchs, J., Kim, M., Petit, J.-E., and Favez, O.: Meteorology-Driven Variability of Air Pollution (PM 1) Revealed with Explainable Machine Learning, Atmospheric Chemistry and Physics, 21, 3919–3948, https://doi.org/10.5194/acp-21-3919-2021, 2021.

Sundararajan, M. and Najmi, A.: The many Shapley values for model explanation, in: International conference on machine learning, pp. 9269–9278, PMLR, 2020.

Takahashi, T., Sugimoto, S., Kawano, T., and Suzuki, K.: Microphysical Structure and Lightning Initiation in Hokuriku Winter Clouds, 124, 13 156–13 181, https://doi.org/10.1029/2018JD030227, 2019.

Takeuti, T., Nakano, M., Brook, M., Raymond, D. J., and Krehbiel, P.: The anomalous winter thunderstorms of the Hokuriku Coast, 83, 2385–2394, https://doi.org/10.1029/JC083iC05p02385, 1978.

Tippett, M. K., Lepore, C., Koshak, W. J., Chronis, T., and Vant-Hull, B.: Performance of a Simple Reanalysis Proxy for U.S. Cloud-to-Ground Lightning, International Journal of Climatology, 39, 3932–3946, https://doi.org/10.1002/joc.6049, 2019.

Toms, B. A., Barnes, E. A., and Hurrell, J. W.: Assessing Decadal Predictability in an Earth-System Model Using Explainable Neural Networks, Geophysical Research Letters, 48, e2021GL093 842, https://doi.org/10.1029/2021GL093842, 2021.

Tost, H., Jöckel, P., and Lelieveld, J.: Lightning and Convection Parameterisations – Uncertainties in Global Modelling, Atmospheric Chemistry and Physics, 7, 4553–4568, https://doi.org/10.5194/acp-7-4553-2007, 2007.

Ukkonen, P. and Mäkelä, A.: Evaluation of Machine Learning Classifiers for Predicting Deep Convection, J. Adv. Model. Earth Sy., 11, 1784–1802, https://doi.org/10.1029/2018MS001561, 2019.

Ukkonen, P., Manzato, A., and Mäkelä, A.: Evaluation of Thunderstorm Predictors for Finland Using Reanalyses and Neural Networks, ournal of Applied Meteorology and Climatology, 56, 2335–2352, https://doi.org/10.1175/JAMC-D-16-0361.1, 2017.

van der Velden, B. H., Kuijf, H. J., Gilhuijs, K. G., and Viergever, M. A.: Explainable artificial intelligence (XAI) in deep learning-based medical image analysis, Medical Image Analysis, 79, 102 470, https://doi.org/10.1016/j.media.2022.102470, 2022.

Wang, D., Zheng, D., Wu, T., and Takagi, N.: Winter Positive Cloud-to-Ground Lightning Flashes Observed by LMA in Japan, 16, 402–411, https://doi.org/10.1002/tee.23310, 2021.

Wang, P., Yuval, J., and O'Gorman, P. A.: Non-local parameterization of atmospheric subgrid processes with neural networks, Journal of Advances in Modeling Earth Systems, p. e2022MS002984, https://doi.org/10.1029/2022MS002984, 2022.

Wood, S. N.: Generalized Additive Models: An Introduction with R, Texts in Statistical Science, Chapman & Hall/CRC, Boca Raton, 2nd edn., https://doi.org/10.1201/9781420010404, 2017.

Wood, S. N., Li, Z., Shaddick, G., and Augustin, N. H.: Generalized Additive Models for Gigadata: Modeling the U.K. Black Smoke Network Daily Data, Journal of the American Statistical Association, 112, 1199–1210, https://doi.org/10.1080/01621459.2016.1195744, 2017.

Zhang, Y., Tiňo, P., Leonardis, A., and Tang, K.: A Survey on Neural Network Interpretability, IEEE Transactions on Emerging Topics in Computational Intelligence, 5, 726–742, https://doi.org/10.1109/TETCI.2021.3100641, 2021.