Amplification of Annual and Diurnal Cycles of Alpine Lightning

Thorsten Simon (✉ thorsten.simon@uibk.ac.at)

University of Innsbruck: Universitat Innsbruck <https://orcid.org/0000-0002-3778-7738>

Georg J. Mayr

University of Innsbruck: Universitat Innsbruck

Deborah Morgenstern

University of Innsbruck: Universitat Innsbruck

Nikolaus Umlauf

University of Innsbruck: Universitat Innsbruck

Achim Zeileis

University of Innsbruck: Universitat Innsbruck

Research Article

Keywords: Lightning location system, ERA5, climate change, generalized additive models, machine learning, complex terrain

Posted Date: October 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-965951/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Amplification of Annual and Diurnal Cycles of Alpine Lightning

Thorsten Simon^{1,2*}, Georg J. Mayr^{2†}, Deborah
Morgenstern^{1,2†}, Nikolaus Umlauf^{1†} and Achim Zeileis^{1†}

^{1*}Department of Statistics, Universität Innsbruck,
Universitätsstrasse 15, Innsbruck, 6020, Austria.

²Department of Atmospheric and Cryospheric Sciences,
Universität Innsbruck, Innrain 52, Innsbruck, 6020, Austria.

*Corresponding author(s). E-mail(s):

Thorsten.Simon@uibk.ac.at, ORCID: 0000-0002-3778-7738;

Contributing authors: Georg.Mayr@uibk.ac.at, ORCID:
0000-0001-6661-9453; Deborah.Morgenstern@uibk.ac.at, ORCID:
0000-0002-0915-5725; Nikolaus.Umlauf@uibk.ac.at, ORCID:
0000-0003-2160-9803; Achim.Zeileis@R-project.org, ORCID:
0000-0003-0918-3766;

[†]These authors contributed equally to this work.

Abstract

The response of lightning to a changing climate is not fully understood. Historic trends of proxies known for fostering convective environments suggest an increase of lightning over large parts of Europe. Since lightning results from the interaction of processes on many scales, as many of these processes as possible must be considered for a comprehensive answer. Recent achievements of decade-long seamless lightning measurements and hourly reanalyses of atmospheric conditions including cloud micro-physics combined with flexible regression techniques have made a reliable reconstruction of lightning down to its seasonally varying diurnal cycle feasible. To include a large variety of land-cover, topographical and atmospheric circulation conditions, the European Eastern Alps and their surroundings are our reconstruction region over a period of four decades. The most intense changes occurred over the high Alps where lightning

activity doubled in the past decade compared to the 1980s. There, the lightning season reaches a higher maximum and starts one month earlier, likely due to the earlier snow melt. Diurnally, the peak is up to 50% stronger with more lightning strikes in the afternoon and evening hours. Signals along the southern and northern alpine rim are similar but weaker whereas the flatlands north of the Alps have no significant trend.

Keywords: Lightning location system, ERA5, climate change, generalized additive models, machine learning, complex terrain

1 Introduction

Lightning has recently been added as an essential climate variable of the Global Climate Observing System (Aich et al, 2018). Cloud-to-ground lightning strikes may damage equipment and structures such as wind turbines (Becerra et al, 2018) and power lines (Cummins et al, 1998), start fires (Reineking et al, 2010) and injure or kill people (Ritenour et al, 2008; Holle, 2016). Further, lightning contributes NO_x and ozone as air pollutants (DeCaria et al, 2005), and lightning threatens permafrost (Finney, 2021). In alpine regions convective events associated with lightning trigger debris flows (Underwood et al, 2016).

Since lightning is an essential climate variable, there is an urgent need to understand how lightning has been evolving over the past decades. This is challenging as lightning is affected at various temporal and spatial scales from micro physics (Houze, 2014) over meso-scale dynamics (Feldmann et al, 2021) to synoptics (Posch, 2018). These scales are all subject to different changes by a changing climate.

We aim at investigating the historic evolution of lightning for the European Eastern Alps. Climatologies show that this region is a hot spot within Europe (Feudale et al, 2013; Wapler, 2013). The interaction of complex topography of the Alps and atmospheric processes such as circulation and radiation leads to persistent forcings (Bertram and Mayr, 2004). Orographic lifting, thermally

induced circulations (plains–mountains, slope winds, valley winds), and lee effects can all trigger lightning in this region (Houze, 2012). Further, this region is exposed to climate change in a special way as the area covered by snow and ice is decreasing (Matiu et al, 2021). A trend analysis of the mean annual number of days with thunderstorms since 1979 indicate an increase over the Alps (Taszarek et al, 2019).

Leveraging recent advancements in three fields enables us to tackle the reconstruction of lightning over a 40 year time period: (1) Decade-long homogeneously detected *lightning observations* (Cummins and Murphy, 2009; Schulz et al, 2016). (2) Physically-based numeric *model outputs* for hourly reanalysis of the atmospheric conditions including micro-physics over several decades (Hersbach et al, 2020). (3) *Flexible regression techniques* from statistics and machine learning (see Hastie et al, 2009, for an overview) that allow to select and link the relevant model outputs to the lightning observations in order to obtain reconstructions. Note that all three elements have to come together to yield a high-quality answer. Just using detected lightning observations over a decade would be too short to assess climate. The physical models alone do not resolve thunderstorms. And the statistical/machine learning techniques are needed for obtaining a satisfactory regression fit with good out-of-sample predictive performance.

Previous studies applied proxies for convection to assess its historic behavior from atmospheric reanalyses. Classically, cloud top height (Price and Rind, 1992), the product of CAPE and precipitation (Romps et al, 2018), the square root of mid-level CAPE multiplied by deep layer shear (Taszarek et al, 2019), CAPE exceeding a threshold conditioned on the occurrence of convective precipitation (Taszarek et al, 2021), or vertical iceflux (Finney et al, 2014) have been applied. However, different proxies or parameterizations lead to contrary

4 *Amplification of Alpine Lightning*

results in the light of a changing climate. For instance, using only iceflux could project a decrease of future lightning (Finney et al, 2018), whereas the product of CAPE and precipitation projected an increase of lightning under climate change (Romps et al, 2014). Further, discussion on this controversy can be found in Murray (2018) and Taszarek et al (2021).

We utilize a flexible regression approach to link the detected lightning observations to the model output from the atmospheric reanalyses. In a first step, a generalized additive model (GAM, Wood, 2017) is trained on the time period where both data sources are overlapping. The possibly nonlinear effects of the atmospheric model outputs on the occurrence of lightning as well as the corresponding annual and diurnal cycles are captured automatically by the GAM using machine learning techniques. In a second step, predictions (or fitted values) from the GAM are obtained for the full time period where the atmospheric model output data is available. Thus, this yields lightning reconstructions for the time period where atmospheric model outputs but no lightning detection observations are available.

Similarly, statistical postprocessing of output from numerical weather prediction (NWP) models using historical observation records is widely used in the weather forecasting literature and termed *model output statistics* (MOS). Therefore we call our approach *lightning MOS* (L-MOS). While L-MOS serves here to reconstruct historic lightning, it could also be used for future projections.

Moreover, other statistical/machine learning approaches have also been used in previous lightning studies. In particular, Ukkonen and Mäkelä (2019) have evaluate various machine learning classifiers for linking reanalysis output to the occurrence of lightning. Although they have not computed reconstructions with their classifiers, they encouraged using this approach to study

climate trends. Beyond lightning, GAMs have proven their capabilities for severe weather trends (Rädler et al, 2018). Similar machine learning approaches have been used to generate wildfire danger maps (Vitolo et al, 2020; Coughlan et al, 2021). Also the field of postprocessing numerical weather predictions leverages these approaches. Probabilistic thunderstorm forecasts generated by statistical NWP postprocessing have a value for air-traffic control (Brunet et al, 2019). Here, the applied techniques range from GAMs combined with objective variable selection (Simon et al, 2018) to deep neural networks (Kamangir et al, 2020).

Our L-MOS reconstructs 40 years (1979–2019) of consistent lightning data for the European Eastern Alps. In particular, it yields the probability for the occurrence of lightning in an ERA5 grid cell within an hour. With this data on hand we are able to tackle the following scientific objectives. We analyze spatio-temporal variability over the last 40 years with emphasis on three subdomains which are associated with different atmospheric processes triggering lightning, namely the northern Alpine rim (NAR), the high Alps (HIA), and the southern Alpine rim (SAR). Further, we identify potential shifts and/or expansions of diurnal cycles and the thunderstorm season. Finally, we investigate climate trends over the past 40 years.

The remainder of the paper is structured as follows. The next section introduces the lightning detection data (Sect. 2.1) and the atmospheric reanalyses (Sect. 2.2). The data are used to build a baseline climatology (Sect. 3.1) and the L-MOS (Sect. 3.2). Sect. 4.1 looks at the crucial variables identified with L-MOS and Sect. 4.2 evaluates its performance. The key results are the reconstructed diurnal and annual cycles (Sect. 4.3) as well as the climate trends of lightning (Sect. 4.4). Finally, we discuss our approach (Sect. 5) and conclude the study (Sect. 6).

Fig. 1 Topography of study area. The data are aggregated to a $32\text{ km} \times 32\text{ km}$ mesh. Colored grids indicate the subdomains. dark blue: Northern Alpine rim (NAR), light blue: Southern Alpine rim (SAR), green: High Alps (HIA).

2 Data

Two data sets are the ingredients for this study. First, the observational data from the lightning location system ALDIS (Sect. 2.1), which serves as target in the baseline climatology and the L-MOS. Second, direct and derived ERA5 parameters (Sect. 2.2) fill the pool of explanatory variables for the L-MOS.

2.1 Lightning Detection Data

The Austrian Lightning Detection & Information System (ALDIS) is part of the European effort EUCLID (Schulz et al, 2016). Detected cloud-to-ground flashes ($> 15\text{ kA}$ and $< -2\text{ kA}$ current) are hourly aggregated to a grid with 32 km spatial mesh. Spatially, the grid ranges from 8.5° E to 16.9° E and from

45.9° N to 49.8° N (Fig. 1). Temporally, data for the boreal summers (Apr–
 Sep) from 2010–2019 are available. For this period of time the lightning data
 was homogeneously detected and processed.

2.2 Atmospheric Reanalysis Output

ECMWF’s fifth reanalysis (ERA5, [Hersbach et al, 2020](#)) comes with a res-
 olution of 32 km horizontally and of 1 hour in time. Further, their vertical
 discretization with 137 levels resolves many details valuable for the descrip-
 tion of convective processes. We pre-select 40 single level variables and extend
 these by deriving another 45 variables from vertical profiles within the tropo-
 sphere, for which we used 74 model levels between the surface (level 137) and
 approximately 15 km altitude (or 120 hPa at level 64) ¹.

The focus of the pre-selection of single level variables and the derivation of
 further variables from model levels is on covering processes relevant for convec-
 tion and lightning from atmospheric environments favorable for the formation
 of thunderstorms to electrical charge separation. For instance, we derived cloud
 top height, wind shear within and below the clouds, the amount of liquid
 and solid water between predefined isotherms, and vertical iceflux in the mid
 troposphere. The data between 2010 and 2019 are used for training and eval-
 uation of the L-MOS, data from 1979 onward are used for the probabilistic
 reconstruction of lightning occurrences.

3 Methods

Our statistical techniques aim at explaining the occurrence of lightning in a
 probabilistic manner. The occurrence of lightning varies over space and time.
 These spatio-temporal variations can be represented by the setting in which

¹L137 model level definitions: [https://www.ecmwf.int/en/forecasts/
 documentation-and-support/137-model-levels](https://www.ecmwf.int/en/forecasts/documentation-and-support/137-model-levels)

they occur, such as longitude, latitude, topography, yearday and time of day. A GAM linking these parameters to lightning serves as baseline climatology (Sect. 3.1). This baseline climatology is extended by parameters from ERA5 to connect the occurrence of lightning to atmospheric processes. This extension is denoted as *lightning MOS* (L-MOS, Sect. 3.2) and is applied for the probabilistic reconstruction of lightning.

Generalized additive models (GAM, Wood, 2017) are the framework for statistical modeling. GAMs are capable of identifying smooth and non-linear effects resulting from the setting (Simon et al, 2017) as well as atmospheric processes (Simon et al, 2018).

3.1 Baseline Climatology

The baseline climatology links the probability $\pi_{baseline}$ of lightning occurrence to the spatio-temporal setting using covariates for geographic space (longitude `lon` and latitude `lat`), time (day of the year `yday` and hour of day `hour`) and ERA5 model topography `topo`. For this purpose we define the following GAM,

$$\text{logit}(\pi_{baseline}) = \beta_0 + f_1(\text{lon}, \text{lat}, \text{hour}) + f_2(\text{yday}) + f_3(\text{topo}). \quad (1)$$

The `logit()` function maps the probability of lightning to the real line. The right hand side is composed of the intercept β_0 and three potentially non-linear functions $f_1()$, $f_2()$ and $f_3()$. The first is a three dimensional smooth effect that allows variations of the diurnal cycle over geographical space. The second and the third add an annual cycle and an effect of the topography.

The non-linear functions f_* () are set up using P-splines and thin plate regression splines (Wood, 2017). The GAM is optimized using restricted maximum likelihood method (Wood et al, 2017). Simon et al (2017) presents details on the usage of GAMs for lightning climatologies.

3.2 Lightning Model Output Statistics

To build the lightning model output statistics (L-MOS), the baseline climatology is extended additively by non-linear effects of ERA5 parameters,

$$\text{logit}(\pi_{\text{era5}}) = \beta_0 + f_1(\text{lon}, \text{lat}, \text{hour}) + f_2(\text{yday}) + f_3(\text{topo}) + g_1(\mathbf{x}_1) + \dots + g_p(\mathbf{x}_p), \quad (2)$$

where $\mathbf{x}_*$ can be replaced by any ERA5 parameter and g_* () is the corresponding non-linear function.

The most important ERA5 parameters are selected objectively from the candidate pool of single level parameters and parameters derived from model level data (Sect. 2.2). The algorithm for the objective selection of ERA5 parameters is a combination of gradient boosting (Bühlmann and Hothorn, 2007) and stability selection (Meinshausen and Bühlmann, 2010). Simon et al (2018) applied the same algorithm for a selection of NWP parameters for a thunderstorm forecasting scheme.

Note that the functional form of the nonlinear effects f_* () in equation 2 need not be the same as in equation 1. They might differ from each other if the non-linear effects of the ERA5 parameters explain parts of the variation that was covered by the effects f_* () in the baseline climatology.

3.3 Reference Models

To evaluate our L-MOS, we pick two popular proxies for convective environments and convert them into binary lightning yes or no variables. The two

proxies are CAPE given convective precipitation, which mimics a situation in which CAPE is released (Taszarek et al, 2021), and iceflux at 450 hPa, which mimics charge separation (Finney et al, 2014). We convert them into binary lightning yes or no variables by setting thresholds. If the threshold is exceeded the variable indicates the occurrence of lightning. The thresholds are computed such that the binary variables have the same average as the observations. For CAPE released the threshold is $260 J kg^{-1}$, and for iceflux the threshold is $1.65 \times 10^{-5} kg_{ice} m_{cloud}^{-2} s^{-1}$. Since these thresholds are calibrated to our domain, they will likely differ from the values used in other studies.

4 Results

Four different topics of results are presented in detail. First, we look at the objectively selected ERA5 parameters for the L-MOS (Sect. 4.1). Second, we compare the performance of the L-MOS against the baseline climatology, and we validate that the reconstructed diurnal cycles follow the observed diurnal cycles (Sect. 4.2). Third, we investigate the evolution of diurnal and annual cycles over the past decades (Sect. 4.3). Fourth, we analyze the climate trends and their spatial distribution (Sect. 4.4).

4.1 Selected Variables

An objective selection algorithm (Sect. 3.2) is applied to extend the climatology with ERA5 parameters. The set of ERA5 parameters consists of two parts (Tab. 1). Single level parameters that can be directly retrieved and parameters that were derived from ERA5 model level data. Selected single level variables are CAPE (`cape`), convective precipitation (`cp`), instantaneous surface sensible heat flux (`ishf`), mid-level cloud cover (`mcc`), two meter temperature (`2t`) and total column supercooled liquid water (`tcslw`). Selected model-level derived

Type	Abbreviation	Description	
ERA5 single level variables	cape	Convective available potential energy.	461
	cp	Convective precipitation.	462
	ishf	Instantaneous surface sensible heat flux.	463
	mcc	Medium cloud cover.	464
	2t	2 meter temperature.	465
Derived variables	tcslw	Total column supercooled liquid water.	466
	cswc2040	Mass of specific snow water content between the -20°C and -40°C isotherms.	467
	cth	Cloud top height in height above ground.	468
	wvc1020	Mass of water vapor between the -10°C and -20°C isotherms.	469

Table 1 ERA5 variables included in the L-MOS. They are automatically selected by the statistical learning algorithm (Sect. 3.2).

variables are the mass of specific snow water content between the -20°C and -40°C isotherms (**cswc2040**), the cloud top height in height above ground (**cth**), and the mass of water vapor between the -10°C and -20°C isotherms (**wvc1020**).

4.2 Performance and Validation

We assess the performance of the L-MOS by computing numerical measures and compare them to the same measures obtained for the baseline climatology. The L-MOS and the baseline climatology fitted to the whole data of ten years explain 33.5 % and 12.4 % of deviance, respectively. Further, we computed the receiver operating characteristics (ROC) in a ten-fold cross-validation experiment. The area under the ROC curve (AUC) measures the ability of the two methods to discriminate between the target classes *lightning* and *no lightning*. The AUC is evaluated for each hour of day, month and region separately. The median AUC (5% quantile, 95% quantile) over these values is 0.90 (0.81, 0.98) for the L-MOS, which outperforms the baseline climatology with 0.58 (0.38, 0.80) by far.

To further validate the L-MOS, we compare the reconstructed probabilities for the period of 2010–2019 with the dichotomous observations (Fig. 2). Again a ten-fold cross-validation is applied. For the comparison, both the fitted and

Fig. 2 Validation of reconstructed diurnal cycles over the period between 2010–2019. The reconstructions are obtained by cross-validation. The dashed lines represent diurnal cycles computed from proxies as discussed in Sect. 5. Columns present diurnal cycles for each month. Rows for each region.

observed values are aggregated to diurnal cycles for each month and subdomain (northern Alpine rim, high Alps and southern Alpine rim). The reconstruction (red solid lines in Fig. 2) closely matches amplitude and phase of the observed diurnal lightning cycle (black solid lines).

Cycles reconstructed with the simple proxies of CAPE released and iceflux (light blue and purple dashed lines in Fig. 2), on the other, deviate strongly from observations. The diurnal cycles obtained from CAPE released peak about 2 to 3 hours too early. Moreover, the maximum probabilities of this proxy over the high Alps are far too high. The diurnal cycles of the proxy derived from iceflux are not as pronounced as the diurnal cycles of the observations. The probabilities during night and morning are too high and the probabilities during afternoon and evening are too low.

Fig. 3 Reconstructed diurnal cycles of probabilities for lightning events averaged over the four decades from 1980s–2010s (color coded). Columns present diurnal cycles for each month. Rows for each region.

4.3 Reconstructed Diurnal and Annual Cycles

Since L-MOS replicates lightning observations so well, it will be used to reconstruct lightning prior to the beginning of homogeneous lightning data. To investigate the diurnal cycles we averaged the reconstructed probabilities for each hour, month and subdomain and display the results by decade in Fig. 3. The 1980s are in blue, the 1990s in green, the 2000s in yellow, and the 2010s in red. The curves for the annual cycle in (Fig. 4) are obtained from aggregating the probabilities for each day of the year and subdomain and further temporal smoothing to account for the strong day-to-day variability of lightning (Simon et al, 2017). In the following we discuss the results for the three subdomains going from North to South.

In the northern Alpine rim, the strongest change of the diurnal cycles (top row in Fig. 3) is found in June, for which the lightning activity between

18 UTC–06 UTC ² increased by about 50% after 2000 compared to the time before. The relative changes for the same periods in May are on the order of 25%. In August we see some decadal variability with the 1990s and 2000s having been less active than the 1980s and the 2010s. These changes in the diurnal cycles affect the annual cycle for this region (left panel in Fig. 4). The lightning season begins earlier after 2000. The second half of the lightning season has decadal variability but no clear trend.

In the high Alps, the diurnal cycles of lightning changed strongly between May–Aug (middle row in Fig. 3). The increase is most prominent in June, for which the probabilities doubled. Also in May the increase is close to doubling. In July and August the increase is also pronounced but the relative change is not as strong as in May and June. The resulting disparate annual cycles (center panel in Fig. 4) can be summarized as follows. In 2010s the season starts earlier and lasts longer. The smoothed annual cycle crosses the 2.5% mark 30 days earlier in the 2010s than in the 1980s. Also the seasonal peak is reached earlier and has an amplitude of more than one percentage point higher than in 1980s and 1990s.

At the southern Alpine rim, lightning activity increases throughout the investigated season from May until September (bottom row in Fig. 3). Over the whole season relative changes of around 50% are not exceptional. As it was the case in the other subdomains, strong changes especially take place after the diurnal peak that extend lightning activity towards the evening and night hours. The annual cycle reconstructed for the southern Alpine rim (right panel in Fig. 4) exhibits that the lightning season in the 2010s is stronger with an earlier onset and extended tail compared to the 1980s.

²During Apr–Sep the European Eastern Alps fall into the Central European Time which is UTC + 2. Thus, a value at 18 UTC refers to 8 pm local time.

Fig. 4 Reconstructed annual cycles of probabilities for lightning events averaged over the four decades from 1980s–2010s (color coded). The light curves in the background are aggregations to the day of the year. The dark curves in the foreground are smoothed versions of the light curves.

Fig. 5 Linear climate trends: Color luminances give the slope per decade of a linear regression for mean probability of lightning within an hour in percent. Desaturated colors indicate trends that are not significant at the 5% level. Only June and afternoons between 13 and 19 UTC are considered.

4.4 Climate Trends

The analysis of diurnal and annual cycles reveals a tremendous increase of lightning activity especially in late spring and early summer. Therefore, further investigation of trends over the period 1979–2019 is performed. Here, we focus on afternoon lightning between 13 UTC–19 UTC as this time of the day exhibits the largest absolute values for lightning probability in the whole domain. To investigate the trends, the reconstructed lighting probabilities for June afternoons are averaged for each year. Then the evolution of the aggregated probabilities over time is explained by a linear regression. The slope coefficient of the linear regression serves as a measure for the trend. First, we present the trends on the scale of the 32 km grid boxes (Fig. 5), and second, on the scale of the three subdomains (Fig. 6).

On the grid box scale no negative trends have been detected in the whole domain. In particular, almost all grids south of 47.5° N show significant increase at the 5% level. Only at the very southeast where the topography flattens into plains the lightning trends are lower and insignificant. Another remarkable outcome is that more than a quarter of cells exceed an increase of one percentage point per decade.

The shape of the region with significant trends matches well the regions with complex topography. In particular the high Alps are witnessing large increases. The strongest trends reaching up to 2.0 percentage points per decade are found on the southern foothills in Italy. Another local hotspot is the region in the very east of the Alps between 14° E–16° E and around 47° N. Finally, we see less strong but still significant positive trends North of the Alps in the Bavarian-Bohemian Forest around 14° E and 48.5° N.

Aggregating the probabilities on the scale of the subdomains (Fig. 6) strengthens the results of the trend analysis. In the high Alps the increase is highest with 1.2 percentage points per decade followed by the southern Alpine

Fig. 6 Linear trends for June afternoons (13 UTC–19 UTC) aggregated for the three subdomains.

rim (0.77) and the northern Alpine rim (0.44). Though the trends for all three subdomains are significant at the 5% level, it is remarkable that the trend in the high Alps explains around third (34.7%) of the variance in this analysis.

5 Discussion

The lightning-MOS approach successfully replicates diurnal and seasonal cycles of observed lightning and reconstructs lightning for pre-measurement periods. Three recent advances contribute to its success: Long enough high-quality and homogeneous lightning measurements, physically consistent atmospheric conditions extending prior to lightning records, and statistical/machine learning techniques to weave the two together, correct the weaknesses of the atmospheric reanalysis and apply their good out-of-sample predictive skill to extend lightning records into the past.

Lightning strikes are rare events. Even at the peak of convective season and time of day, probabilities for the occurrence linger mostly in the single digits (cf. Figs. 2–4) and spatio-temporal variation are high. However, the additive nature of GAMs allows to add information from grid cells that share spatial

characteristics (location, altitude), temporal characteristics (diurnal and seasonal), and the state of the atmosphere to achieve a successful replication of the lightning measurements. GAMs allow a *functional* dependence on all these variables. The shape of these functions is directly derived from the data and thus potentially non-linear. This flexibility permits to properly account for the diurnal cycle of lightning, which simple, single proxy variables (Finney et al, 2014; Taszarek et al, 2021, among others) cannot accomplish. This flexibility is also sufficient to achieve good out-of-sample performance with “only” a decade’s worth of lightning measurements as the cross-validation (Sect. 4.2) demonstrates. The additive and functional nature of GAMs makes it possible to incorporate approximations of the effects of multiple processes known to lead to lightning that can be derived from the available location, temporal and atmospheric state information—including favorable environments, triggers and electrical charge separation.

GAMs also compensate for weaknesses of the observational data. They allow to extend the relatively short duration of lightning measurements, derive a phenomenon that is not resolved in the reanalysis data, and correct the somewhat misrepresented diurnal cycle of convection in ERA5. As the blue curves for CAPE conditional on convective precipitation in Fig. 2 show, convection in ERA5 has a premature peak and too strong an amplitude, a result that Watters et al (2021) also found. Yet with a GAM (red curves in Fig. 2), phase and amplitude become well reproduced.

Our L-MOS suggests that the increase in lightning activity over the high Alps is connected to the decrease of spring snow cover (Matiu et al, 2021). The processes captured by L-MOS include favorable environments for thunderstorms (`cape` and `cp`), charge separation (`cswc2040` and `wvc1020`), and triggers (`ishf`). Sensible heat flux as a trigger is of special interest in the Alps.

As spring snow covers have been decreasing over the past decades ([Matiu et al, 2021](#)), thunderstorms can be thermally induced more often from energy originating from sensible heat fluxes.

6 Conclusion

Lightning in the European Eastern Alps is probabilistically reconstructed back to 1979 by building a L-MOS that takes reanalysis output (ERA5) to explain the occurrence of lightning. The hourly resolution of ERA5 allows us to reconstruct and investigate diurnal cycles of thunderstorms. Further, the high vertical resolution of cloud micro physics makes it possible to incorporate a wide range of atmospheric processes into the L-MOS. The validation of the reconstructed probabilities against observations from ALDIS shows a good performance of the method. The reconstruction over 40 years (1979–2019) reveals amplification of the diurnal and annual cycles. This is in particular the case in spring in the high Alps where lightning activity doubled, and at the southern Alpine rim throughout the season. A trend analysis reveals that the strongest trend is found in the high Alps. This sensible region is prone to climate change. In spring a strong decrease in snow cover has been observed over the past decades ([Matiu et al, 2021](#)). The decrease in snow cover affects the heat exchange between atmosphere and surface, and thus could also influence triggering mechanisms for convection and lightning.

Data Availability

The ERA5 data are accessible via the Copernicus Climate Change Service (C3S) Climate Data Store. ALDIS data are available on request from OVE-ALDIS (aldis@ove.at)—fees may be charged.

Computational Details

The results in this study were achieved using R, a software environment for statistical computing and graphics. The add-on packages **mgcv** and **bam** (Umlauf et al, 2021) were used for statistical modeling. We used the netCDF4 data format.

Statements and Declarations

The authors declare that they have no conflict of interest.

Acknowledgments

This work was funded by the Austrian Science Fund (FWF, grant no. P 31836) and the Austrian Research Promotion Agency (FFG, grant no. 872656). The computational results presented here have been achieved (in part) using LEO HPC infrastructure of Universität Innsbruck. We thank Gerhard Diendorfer and Wolfgang Schulz from OVE-ALDIS for data support. ERA5 was downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store.

References

- Aich V, Holzworth R, Goodman SJ, et al (2018) Lightning: A new essential climate variable. Eos 99. <https://doi.org/10.1029/2018EO104583>, published on 07 September 2018
- Becerra M, Long M, Schulz W, et al (2018) 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>

- Bertram I, Mayr GJ (2004) Lightning in the Eastern Alps 1993–1999, part I: Thunderstorm tracks. *Nat Hazards Earth Syst Sci* 4(4):501–511. <https://doi.org/10.5194/nhess-4-501-2004>
- Brunet D, Sills D, Driedger N (2019) On the evaluation of probabilistic thunderstorm forecasts and the automated generation of thunderstorm threat areas during Environment Canada Pan Am Science Showcase. *Weather and Forecasting* 34(5):1295–1319. <https://doi.org/10.1175/waf-d-19-0011.1>
- Bühlmann P, Hothorn T (2007) Boosting algorithms: Regularization, prediction and model fitting. *Statist Sci* 22(4):477–505. <https://doi.org/10.1214/07-STS242>
- Coughlan R, Di Giuseppe F, Vitolo C, et al (2021) Using machine learning to predict fire-ignition occurrences from lightning forecasts. *Meteorological Applications* 28(1):e1973. <https://doi.org/10.1002/met.1973>
- Cummins K, Krider E, Malone M (1998) The US National Lightning Detection Network and applications of cloud-to-ground lightning data by electric power utilities. *IEEE Transactions on Electromagnetic Compatibility* 40(4):465–480. <https://doi.org/10.1109/15.736207>
- Cummins KL, Murphy MJ (2009) An overview of lightning locating systems: History, techniques, and data uses, with an in-depth look at the U.S. NLDN. *IEEE Transactions on Electromagnetic Compatibility* 51(3):499–518. <https://doi.org/10.1109/TEM.2009.2023450>
- DeCaria AJ, Pickering KE, Stenchikov GL, et al (2005) 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(D14). <https://doi.org/10.1029/2004JD005556>
- Feldmann M, Germann U, Gabella M, et al (2021) A characterisation of Alpine mesocyclone occurrence. *Weather and Climate Dynamics Discussions* 2021:1–26. <https://doi.org/10.5194/wcd-2021-53>
- Feudale L, Manzato A, Micheletti S (2013) A cloud-to-ground lightning climatology for north-eastern Italy. *Adv Sci Res* 10(1):77–84. <https://doi.org/10.5194/asr-10-77-2013>
- Finney DL (2021) Lightning threatens permafrost. *Nature Climate Change* <https://doi.org/10.1038/s41558-021-01016-7>
- Finney DL, Doherty RM, Wild O, et al (2014) Using cloud ice flux to parametrise large-scale lightning. *Atmospheric Chemistry and Physics* 14(23):12,665–12,682. <https://doi.org/10.5194/acp-14-12665-2014>
- Finney DL, Doherty RM, Wild O, et al (2018) A projected decrease in lightning under climate change. *Nature Climate Change* 8(3):210–213. <https://doi.org/10.1038/s41558-018-0072-6>
- Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd edn. Springer Series in Statistics Springer, Berlin, <https://doi.org/10.1007/978-0-387-84858-7>
- Hersbach H, Bell B, Berrisford P, et al (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146(730):1999–2049. <https://doi.org/10.1002/qj.3803>

- Holle RL (2016) A summary of recent national-scale lightning fatality studies. *Weather, Climate, and Society* 8(1):35–42. <https://doi.org/10.1175/WCAS-D-15-0032.1>
- Houze RA (2012) Orographic effects on precipitating clouds. *Rev Geophys* 50(1):1–47. <https://doi.org/10.1029/2011RG000365>
- Houze RA (2014) *Cloud Dynamics, International Geophysics*, vol 104. Academic Press, <https://doi.org/10.1016/B978-0-12-374266-7.00003-2>
- Kamangir H, Collins W, Tissot P, et al (2020) A deep-learning model to predict thunderstorms within 400 km^2 South Texas domains. *Meteorological Applications* 27(2):e1905. <https://doi.org/10.1002/met.1905>
- Matiu M, Crespi A, Bertoldi G, et al (2021) Observed snow depth trends in the European Alps: 1971 to 2019. *The Cryosphere* 15(3):1343–1382. <https://doi.org/10.5194/tc-15-1343-2021>
- Meinshausen N, Bühlmann P (2010) Stability selection. *J Roy Stat Soc B* 72(4):417–473. <https://doi.org/10.1111/j.1467-9868.2010.00740.x>
- Murray LT (2018) An uncertain future for lightning. *Nature Climate Change* 8(3):191–192. <https://doi.org/10.1038/s41558-018-0094-0>
- Posch T (2018) *Weather Patterns Leading to Thunderstorms in the Eastern Alps*. University of Innsbruck (masterthesis), available online under URN: <https://resolver.obvsg.at/urn:nbn:at:at-ubi:1-32693>
- Price C, Rind D (1992) A simple lightning parameterization for calculating global lightning distributions. *Journal of Geophysical Research: Atmospheres* 97(D9):9919–9933. <https://doi.org/10.1029/92JD00719>

- Rädler AT, Groenemeijer P, Faust E, et al (2018) Detecting severe weather trends using an additive regressive convective hazard model (AR-CHaMo). *J Appl Meteor Climatol* 57(3):569–587. <https://doi.org/10.1175/JAMC-D-17-0132.1>
- Reineking B, Weibel P, Conedera M, et al (2010) Environmental determinants of lightning- v. human-induced forest fire ignitions differ in a temperate mountain region of Switzerland. *International Journal of Wildland Fire* 19(5):541–557. <https://doi.org/10.1071/WF08206>
- Ritenour AE, Morton MJ, McManus JG, et al (2008) Lightning injury: A review. *Burns* 34(5):585–594. <https://doi.org/10.1016/j.burns.2007.11.006>
- Romps DM, Seeley JT, Vollaro D, et al (2014) Projected increase in lightning strikes in the United States due to global warming. *Science* 346(6211):851–854. <https://doi.org/10.1126/science.1259100>
- Romps DM, Charn AB, Holzworth RH, et al (2018) CAPE times P explains lightning over land but not the land-ocean contrast. *Geophys Res Lett* 45(22):12,623–12,630. <https://doi.org/10.1029/2018GL080267>
- Schulz W, Diendorfer G, Pedebay S, et al (2016) The European lightning location system EUCLID part 1: Performance analysis and validation. *Natural Hazards and Earth System Sciences* 16(2):595–605. <https://doi.org/10.5194/nhess-16-595-2016>
- Simon T, Umlauf N, Zeileis A, et al (2017) Spatio-temporal modelling of lightning climatologies for complex terrain. *Nat Hazards Earth Syst Sci* 17(3):305–314. <https://doi.org/10.5194/nhess-17-305-2017>

- Simon T, Fabsic P, Mayr GJ, et al (2018) Probabilistic forecasting of thunderstorms in the Eastern Alps. *Mon Wea Rev* 146(9):2999–3009. <https://doi.org/10.1175/MWR-D-17-0366.1>
- Taszarek M, Allen J, Púčik T, et al (2019) A climatology of thunderstorms across Europe from a synthesis of multiple data sources. *J Climate* 32(6):1813–1837. <https://doi.org/10.1175/JCLI-D-18-0372.1>
- Taszarek M, Allen JT, Brooks HE, et al (2021) Differing trends in United States and European severe thunderstorm environments in a warming climate. *Bull Amer Meteor Soc* 102(2):296–322. <https://doi.org/10.1175/BAMS-D-20-0004.1>
- Ukkonen P, Mäkelä A (2019) Evaluation of machine learning classifiers for predicting deep convection. *J Adv Model Earth Sy* 11(6):1784–1802. <https://doi.org/10.1029/2018MS001561>
- Umlauf N, Klein N, Simon T, et al (2021) bamlss: A Lego toolbox for flexible Bayesian regression (and beyond). Tech. rep., accepted for publication in *Journal of Statistical Software*, preprint available on arXiv: <https://arxiv.org/abs/1909.11784>
- Underwood SJ, Schultz MD, Berti M, et al (2016) Atmospheric circulation patterns, cloud-to-ground lightning, and locally intense convective rainfall associated with debris flow initiation in the Dolomite Alps of northeastern Italy. *Natural Hazards and Earth System Sciences* 16(2):509–528. <https://doi.org/10.5194/nhess-16-509-2016>
- Vitolo C, Di Giuseppe F, Barnard C, et al (2020) ERA5-based global meteorological wildfire danger maps. *Nature* 7(1):216. <https://doi.org/10.1038/>

s41597-020-0554-z

Wapler K (2013) High-resolution climatology of lightning characteristics within Central Europe. *Meteor Atmos Phys* 122(3-4):175–184. [https://doi.org/10.](https://doi.org/10.1007/s00703-013-0285-1)

[1007/s00703-013-0285-1](https://doi.org/10.1007/s00703-013-0285-1)

Watters D, Battaglia A, Allan RP (2021) The diurnal cycle of precipitation according to multiple decades of global satellite observations, three CMIP6 models, and the ECMWF reanalysis. *Journal of Climate* 34(12):5063–5080. <https://doi.org/10.1175/JCLI-D-20-0966.1>

Wood SN (2017) *Generalized Additive Models: An Introduction with R*, 2nd edn. Texts in Statistical Science, Chapman & Hall/CRC, Boca Raton

Wood SN, Li Z, Shaddick G, et al (2017) Generalized additive models for gigadata: Modeling the U.K. black smoke network daily data. *Journal of the American Statistical Association* 112(519):1199–1210. <https://doi.org/10.1080/01621459.2016.1195744>