Supporting Information for "Spatio-seasonal risk assessment of upward lightning at tall objects using meteorological reanalysis data"

Isabell Stucke[1,2], Deborah Morgenstern[1,2],Gerhard Diendorfer[3], Georg J. Mayr[2], Hannes Pichler[3], Wolfgang Schulz[3], Thorsten Simon[4], Achim Zeileis[1]

[1]Institute of Statistics, University of Innsbruck, Austria, Innsbruck

[2]Institute of Atmospheric and Cryospheric Sciences, University of Innsbruck, Austria, Innsbruck

[3]OVE Service GmbH, Dept. ALDIS (Austrian Lightning Detection & Information System), Austria, Vienna

[4]Department of Mathematics, University of Innsbruck, Austria, Innsbruck

Contents of this file

Introduction This Supporting Information file contains text sections, a figure, and a table. First, the procedure and equations for calculating the effective height of tall objects are presented. Then the concept of a confusion matrix is explained. Then a figure shows an example of a single decision tree constructed with larger-scale meteorological variables

---

Corresponding author: I. Stucke, Institute of Statistics University of Innsbruck, Innsbruck, Universitätsstrasse 15, 6020, Austria. (isabell.stucke@uibk.ac.at)

available or derived from ERA5. The final table lists the meteorological variables used in the study.

0.1. Estimation of the effective height

The effective height is computed following (Zhou et al., 2010) by assuming a hemispherical mountain:

using:

where H_{eff} (m) is the effective height and h (m) is the actual height of the object. U_{lc} (kV) is the continuous leader inception potential due to the cloud charges, R (m) is a geometrical parameter, a (m) is the mountain height, which in the current study is taken to be the difference between the 1 km^2 mean elevation and the elevation at which the object is located to also account for the surrounding terrain. E_g (kV/m) is the ambient uniform electric field. For more details see (Zhou et al., 2010).

0.2. Understanding a confusion matrix

		Actual	
		Positive	Negative
Predicted	Positive	True positive	False positive
	Negative	False negative	True negative

A true positive rate is the proportion of true positive divided by the sum of true positives and false negatives. The false positive rate on the other hand is the proportion of false positives divided by the sum of true positives and false positives.

0.3. Example of a decision tree

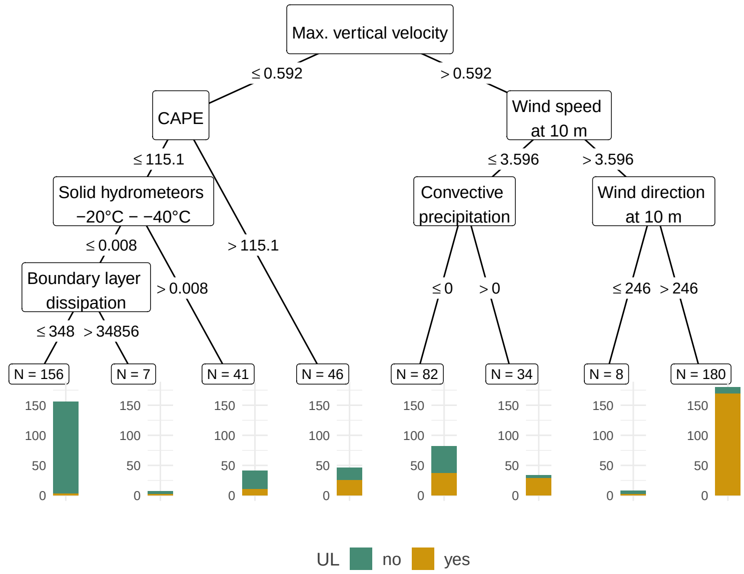

Figure S1. Example of a decision tree. Meteorological variables in the nodes are splitted according to the split points (numbers at the solid lines). Terminal nodes (bars) give the decision. The number of observations included in the decision pars is given above the terminal nodes as N.

Figure S1 shows the structure of a single decision tree. It shows several nodes, each associated with specific split variables. Initially, the maximum large-scale upward velocity serves as the primary split variable. Thresholds between nodes indicate where the split variable is splitted for optimal performance. Following a single UL observation along the path determined by these thresholds leads to a terminal node, represented by the bottom bars. The colors of these bars indicate the number of observations assigned to each terminal node, indicating UL or no UL prediction.

0.4. List of variables included in the random forest models

Table S1. Table of larger-scale variables taken from ERA5 and variables derived from ERA5. The derived variables are suggested to be potentially important in the charging process of a thundercloud or for the development of convection.

Variable	**Unit**	**Variable**	**Unit**
Cloud base height above ground	m agl	Convective precipitation (rain + snow)	m
Large scale precipitation	m	Cloud size	m
Maximum precipitation rate (rain + snow)	kg m^{-2} s^{-1}	Ice crystals (total column, tciw)	kg m^{-2}
Solid hydrometeors (total column, tcsw)	kg m^{-2}	Supercooled liquid water (total column, tcslw)	kg m^{-2}
Water vapor (total column)	kg m^{-2}	Integral of cloud frozen water flux divergence	kg m^{-2} s^{-1}
Vertical transport of liquids around -10 °C	kg Pa s^{-1}	Ice crystals (-10 °C - -20 °C)	kg m^{-2}
Ice crystals (-20 °C - -40 °C)	kg m^{-2}	Cloud water droplets (-10 °C - -20 °C)	kg m^{-2}
Solid hydrometeors (-10 °C - -20 °C)	kg m^{-2}	Solid hydrometeors (-20 °C - -40 °C)	kg m^{-2}
Solids (cswc + ciwc) around -10 °C	kg m^{-2}	Liquids (clwc + crwc) around -10 °C	kg m^{-2}
2 m dew point temperature	K	Mean vertically integrated moisture convergence	kg m^{-2} s^{-1}
Water vapor (-10 °C - -20 °C)	kg m^{-2}	Boundary layer height	m
Surface latent heat flux	J m^{-2}	Surface sensible heat flux	J m^{-2}
Downward surface solar radiation	J m^{-2}	Convective available potential energy	J kg^{-1}
Convective inhibition present	binary	Mean sea level pressure	Pa
Height of -10 °C isotherm	m agl	Boundary layer dissipation	J m^{-2}
Maximum larger-scale upward velocity	Pa s^{-1}	Total cloud shear	m s^{-1}
Wind speed at 10 m	m s^{-1}	Wind direction at 10 m	°
Shear between 10 m and cloud base	m s^{-1}		

References

Zhou, H., Theethayi, N., Diendorfer, G., Thottappillil, R., & Rakov, V. A. (2010). On estimation of the effective eeight of towers on mountaintops in lightning incidence studies. *Journal of Electrostatics*, *68*(5), 415-418. Retrieved from https://www.sciencedirect.com/science/article/pii/S030438861000077X doi: https://doi.org/10.1016/j.elstat.2010.05.014