Modeling and projecting health-relevant combined ozone and temperature events in present and future Central European climate

Sally Jahn[1] · Elke Hertig[2]

Abstract

Statistical models to evaluate the relationships between large-scale meteorological conditions, prevailing air pollution levels and combined ozone and temperature events, were developed during the 1993–2012 period with Central Europe as regional focus. Combined ozone and temperature events were defined based on the high frequency of coinciding, health-relevant elevated levels of daily maximum tropospheric ozone concentrations (based on running 8-h means) and daily maximum temperature values in the peak ozone and temperature season from April to September. By applying two different modeling approaches based on lasso, logistic regression, and multiple linear regression mean air temperatures at 850 hPa, ozone persistence, surface thermal radiation, geopotential heights at 850 hPa, meridional winds at 500 hPa, and relative humidity at 500 hPa were identified as main drivers of combined ozone and temperature events. Statistical downscaling projections until the end of the twenty-first century were assessed by using the output of seven models of the Coupled Model Intercomparison Project Phase 5 (CMIP5). Potential frequency shifts were evaluated by comparing the mid- (2031–2050) and late-century (2081–2100) time windows to the base period (1993–2012). A sharp increase of ozone-temperature events was projected under RCP4.5 and RCP8.5 scenario assumptions with respective multi-model mean changes of 8.94% and 16.84% as well as 13.33% and 37.52% for mid- and late-century European climate.

Keywords Air temperature · Air pollution · Climate change · Human health · Statistical downscaling · Central Europe

Introduction

Air pollution poses the single largest environmental risk to human health in Europe resulting in a substantial public health burden for the European population (EEA 2019). Tropospheric ozone (O_3), representing one major air pollutant, is not directly emitted to the atmosphere, but produced by a photochemical chain reaction in the presence of solar radiation. It is formed by the reaction of precursor gases such as volatile organic compounds (VOC), carbon monoxide (CO), or nitrogen oxides (NO_x) that is enhanced by rising air temperatures and solar radiation.

Elevated levels of O_3 cause a variety of human health effects primary affecting the cardio-pulmonary system (Nuvolone et al. 2018; Srebot et al. 2009; WHO 2006; WHO 2013). The severity and extent of symptoms are highly determined via duration and intensity of the exposure to the reactive and oxidative O_3 gas. Elevated concentrations have irritating effects on eyes, mucous membranes, and airways. Lung inflammation and tissue damage, asthma, a reduction of the self-cleaning mechanism of the bronchi, cardiac arrhythmia, heart attacks, and heart failure are possible resulting respiratory or cardiovascular diseases (Eis et al. 2010; Srebot et al. 2009; WHO 2006). Elevated mortality levels were linked to tropospheric O_3, while potentially susceptible people with, e.g., corresponding previous illnesses show a higher disease and mortality risk (Eis et al. 2010; WHO 2013). The Ambient Air Quality Directive (AQD) of the European Union suggests 120 μg/m^3 (daily maximum 8-h mean) as a target value for

✉ Sally Jahn
sally.jahn@geo.uni-augsburg.de

1 Regional Climate Change and Health, Institute of Geography and Faculty of Medicine, University of Augsburg, Alter Postweg 118, 86159 Augsburg, Germany

2 Regional Climate Change and Health, Faculty of Medicine, University of Augsburg, Alter Postweg 118, 86159 Augsburg, Germany

protecting human health from O_3 (EEA 2019). The WHO even recommended the use of 100 μg/m^3 in their Air Quality Guidelines (AQG), admitting that even under this threshold, some sensitive individuals may suffer from negative health effects (WHO 2006). The WHO (2013) assumed that the implemented target values to protect humans from ozone pollution are all in all too high. Hertig et al. (2019) analyzed the relationship between daily maximum 1-h ozone concentrations and myocardial infarction (MI) frequencies. The city of Augsburg (Bavaria, Germany) was the regional focus. In conclusion, they argued that the existing AQD target value is only suitable to a limited extent as enhanced MI risks already occurred at median to moderately high ozone pollution levels. A maximum risk was found at approximately the 75th percentile with the value referring to 116 μg/m^3.

Elevated air temperature levels (Baccini et al. 2008; Hajat and Kosatky 2010; Song et al. 2017) and heatwave episodes (Anderson and Bell 2011; Gasparrini and Armstrong 2011; Guo et al. 2017; Robine et al. 2012) can negatively affect human health and were associated with increased mortality. A higher cardiovascular, cerebrovascular, or respiratory mortality rate just represents the final extreme end of a variety of adverse impacts on human health. Increasing thermal load can lead to or worsen health effects, for example, severe dehydration, heat exhaustion, cramps, syncope, oedema, thrombogenesis, heat rush, and life-threatening heatstrokes (McGregor et al. 2015). Health impacts are determined by the level of exposure with respect to duration, severity, and frequency as well as the exposed population and its sensitivity (Matthies et al. 2008). There exist a vast number of worldwide as well as national-based indicators and target values to describe extreme temperatures and periods of excessive heat (e.g., DWD 2020; ETCCDI 2009).

As indicated, due to the specific characteristics of ozone formation and further underlying processes, high ozone concentrations often co-occur with elevated air temperature levels (Fiore et al. 2015). As exposure to poor air quality as well as thermal load both already affect human health independently, their combined occurrence poses an even intensified threat to human life, especially as synergistic effects lead to a risk beyond the sum of their individual effects (Katsouyanni and Analitis 2009; WHO 2008). Thus, for example, ozone pollution is an enhanced threat to human health on hot days (Pattenden et al. 2010), while with prolonged periods of heat associated mortality is higher during ozone pollution events (Analitis et al. 2014).

The occurrence of high temperatures and elevated levels of ozone are substantially influenced by meteorological and air pollution conditions. As both health stressors are apparently target variables to assess health burden occasions for the European population and are dependent on recent and future climatic conditions, a variety of assessments exist, investigating separately air temperature or O_3 levels in Europe. Future changes of these health stressors are additionally analyzed in terms of anthropogenic induced global climate change. Recent studies found that precursor emissions show a substantial impact on the relationship between the strongly correlated ozone and temperature target variables (Bloomer et al. 2009; Coates et al. 2016; Sillman and Samson 1995). Meteorological and synoptic conditions influencing O_3 pollution (Carro-Calvo et al. 2017; Otero et al. 2016) or elevated air temperatures on single or consecutive days (Black et al. 2004; Krueger et al. 2015) were evaluated for Europe. Various studies analyzed future health outcomes related to air pollution (Fang et al. 2013; Hendriks et al. 2016) or heat events (Gasparrini et al. 2017; Takahashi et al. 2007) under twenty-first century climate change. A growing number of days with temperature extremes and heat waves were identified in Europe (Jacob et al. 2014; Meehl and Tebaldi 2004; Schoetter et al. 2015). By not considering future anthropogenic mitigation strategies and emission policies, an increase of surface ozone concentrations over Europe in summer was found (Katragkou et al. 2011), largely for high-percentile O_3 levels and polluted environments (Schnell and Prather 2017). Even if future-projected ozone concentration levels are highly determined by changing precursor emission levels, there is growing evidence that the increasing effect of a warming climate could negate recent and future pollution mitigation strategies and policies (Colette et al. 2015; Hendriks et al. 2016). Even with a strong reduction of precursor emissions, ozone standards may still be violated in the future at individual stations or regions (Moghani and Archer 2020).

Only a rare number of studies investigated concurrent elevated temperature and O_3 levels under recent and future climatic conditions. Hertig (2020) assessed the relationship between large-scale meteorological mechanisms and elevated levels of daily maximum ground-level temperature and ozone concentrations for Bavarian cities. Furthermore, combined threshold exceedances of both target variables were analyzed. Projections under RCP8.5 were provided to illustrate changes of these health stressors under future climate change. A strong frequency increase of co-occurring O_3 pollution and thermal load events was identified. A recent study by Meehl et al. (2018) investigated the relationship between heat waves and surface ozone concentrations around the world under RCP6.0 emission scenarios in the twenty-first century. For most areas, a decline was found in ozone concentrations on even intensifying future heat wave days compared to non-heat wave days. In contrast, an increase was assessed for most areas by keeping anthropogenic precursor emissions strongly influencing O_3 pollution levels constant at 2005 levels.

In summary, merely a few assessments exist analyzing the co-occurrence of combined elevated temperature and O_3 levels representing a relevant threat for human health in Europe. Enhanced attention should be paid to the possible impacts of future climate change on these harmful occasions.

Thus, the motivation for this study is to assess the relationships between meteorological conditions and the joint occurrence of elevated tropospheric ozone concentrations and air temperature values over Central Europe. Co-occurring high levels of the two variables are considered to define combined ozone and temperature events (hereafter "o-t-events"; please refer to Table S1 in the Supplementary Material (Online Resource 1) to get an overview of all used abbreviations and acronyms). The response of the o-t-events to meteorological factors as well as prevailing atmospheric and O_3 pollution conditions is investigated to better understand the main drivers of these highly health-relevant events. Projections providing an integration of climate change scenario assumptions until the end of the twenty-first century are presented based on seven models of the fifth phase of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al. 2012). Thus, potential frequency shifts of such health burden occurrences under future climate conditions are assessed.

This paper is organized as follows. The "Data" section presents the initial selection and preprocessing of predictand and predictors by additionally referring to their underlying respective datasets (e.g., reanalyses and earth system model (ESM) output). The "Methods" section introduces all applied statistical model techniques and approaches, the framework of their implementation, the metrics to evaluate their performance to assess the occurrence of combined o-t-events, and the employment of the subsequent projections. The "Results" section presents the results of all statistical modeling and projecting processes. Finally, the discussion and concluding remarks can be found in the "Discussion" and "Conclusion" sections.

Data

Predictand data

Since the main drivers of concurrent elevated air temperature and O_3 levels as well as possible future frequency shifts are assessed, the o-t-events represent the target variable for the statistical downscaling models and subsequent projections. Accordingly, stations with measurement data on surface daily maximum ozone and air temperature in Europe were selected to define these combined events.

Station-based air quality and temperature data

Station-based ozone data were retrieved from the European air quality database data product AirBase version 8 maintained by the European Environment Agency (EEA) (EEA 2014). Data collected within the subsequent Air Quality e-Reporting starting from 2013 based on new rules for reciprocal exchange of information and reporting were disregarded to

assess homogenously submitted and prepared air quality monitoring data and information sets. Valid daily maximum ozone values (selected by EEA's predefined quality flags indicating the quality of the data) based on running 8-h means (henceforth named "MDA8O3") calculated from the corresponding hourly data were retrieved. An observational base period of 20 years from 1993 to 2012 was chosen as temporal focus. Beside the large-scale prevailing meteorological and climatic conditions, location-specific ozone concentrations are influenced by various station environment characteristics. To guarantee the best possible homogeneity of the database, the impact of factors other than meteorological variables, e.g. altitudinal vegetation change, crucially influencing ozone concentrations was minimized for all selected stations. Thus, only ozone stations located primarily in the planar up to the colline zone with an altitude height under 700 m were considered.

Station-based daily maximum surface air temperature (hereafter "TX") observations based on adjusted, homogenized blended temperature series were obtained from the European Climate Assessment & Dataset project (ECA&D) (Klein Tank et al. 2002). Only valid TX observations (selected by ECA&D's predefined quality code) in the country-based blended series of the ECA&D dataset were considered. Only temperature stations in the spatial vicinity of the ozone stations were selected. In order to reproduce the respective temperature as accurately as possible, merely temperature stations located at a distance of at most 10 km and with a maximum altitude difference of 200 m to the corresponding ozone station were chosen.

Station selection

Based on the respective database, all possible station pairs were evaluated to select suitable matches. To minimize the influence of missing values, each ozone and temperature station was required to have at least 75% of valid daily data (valid data for MDA8O3 and TX as defined above). Spearman rank correlation coefficients calculated between the daily and monthly MDA8O3 and TX time series of each station pair show primarily a strong relationship between both variables in spring to late summer. Thus, all analyses are based on the months from April to September, herein after referred to as the ozone-temperature season (hereafter "o-t-season"). Only station pairs with an on average monthly correlation coefficient of at least 0.5 in the o-t-season were selected, leading to an exclusion of eight ozone stations. All of these stations but two (located in Greece) were spread over Northern Europe (namely Great Britain, Finland, and Sweden), showing a weaker correlation between both target variables probably due to the overall, high latitude-based lower solar radiation durations, and intensities. Temperature time series were directly assigned to the paired ozone station to create a combined time series per location. Only station pairs with at least 75% of valid season-

specific data across all years, based on their combined time series, were selected.

As a result, 85 station pairs in Central Europe build the regional focus of the study. The mean distance and the mean altitude difference between all paired stations is 4.72 km (min 0.2, max 9.91, median 4.64) and 31.54 m (min 0, max 146, median 10), respectively. These location differences between TX and MDA8O3 are not further referred to throughout the paper, but should always be kept in mind. Hereafter, "station" is used to refer to these linked ozone and temperature station pairs, with all spatial station information being represented for the respective ozone station. An overview of all finally chosen ozone stations with more detailed station metadata and station pair characteristics is given in Table S2 in the Supplementary Material (Online Resource 1).

To assess the different air quality settings of each location, the finally selected stations are categorized in five station classes based on EEA's predefined types of station and area as follows: urban traffic ("ut", 13 stations), urban background ("ub", 37 stations), suburban background ("sb", 27 stations), rural industrial ("ri", 2 stations), and rural background ("rb", 6 stations). The selected stations are spread across five different countries in Central Europe, namely Austria, Belgium, Germany, The Netherlands, and Switzerland. Throughout the remaining sections, the term "Central Europe" is used to refer to this set of countries. The specific location of each analyzed station together with its respective class, highlighted by shape and color, is shown in Fig. 1. On this and all following maps showing spatial distributions and locations, x- and y-

axis represent longitude and latitude. Apart from that, hereafter, only the shape of the stations indicates the class of each station.

Combined events

Health-relevant, co-occurring high levels of MDA8O3 and TX were used to define o-t-events. Based on a 31-day window, the 75th percentiles for both target variables were calculated with respect to daily mean averages across all years in the base period. As a result, fixed daily values to define threshold exceedances for each variable and day in the o-t-season for all subsequent analysis in the observation, historical, and future projection period were generated. The selection of the 75th percentile was determined by the results of previous health-related studies presented and outlined in the "Introduction" section as well as the actual respective ozone and temperature values the percentiles amounted to taking into account all analyzed stations in Central Europe. More details on this will be given in the "Results" section. Temperature levels as well as ozone concentrations need to exceed these percentile-based thresholds. Additionally, ozone pollution levels surpassing a threshold of 100 $\mu g/m^3$ were also considered as a health-relevant ozone event in accordance with the WHO AQG. Thus, days with thermal load or TX exceedances were observed if TX values rose above the 75th percentile-based thresholds, while MDA8O3 exceedances were defined for days with ozone pollution levels beyond the respective 75th percentile-based or 100 $\mu g/m^3$ thresholds. If not otherwise specified, herein after TX and MDA8O3 exceedances refer to these definitions.

On combined event days, mean and median MDA8O3 and TX levels in the o-t-season amounted to 121.55 $\mu g/m^3$ (min 39.75 $\mu g/m^3$, max 262.73 $\mu g/m^3$) and 119.15 $\mu g/m^3$ as well as 27.5 °C (min 12.4 °C, max 40.2 °C) and 27.9 °C, respectively. Frequency analyses in the base period revealed a higher amount of event days from April to September compared to the remaining months, confirming the defined o-t-season based on spearman rank correlation coefficients.

Predictor data and selection

Predictor metrics belonged to two different types. The first type was considered as "meteorological". Suitable predictors to assess and project o-t-events were chosen and preselected based on a preliminary screening process using reanalysis and ESM data. The second type was considered as "persistence-based" and comprised only O_3 pollution persistence metrics.

The initial selection of predictor variables was determined by literature review (e.g., Carro-Calvo et al. 2017; Hertig 2020; Krueger et al. 2015; Otero et al. 2016), data availability in the chosen ERA5 reanalysis (from the European Centre for Medium-Range Weather Forecasts (ECMWF); Hersbach and
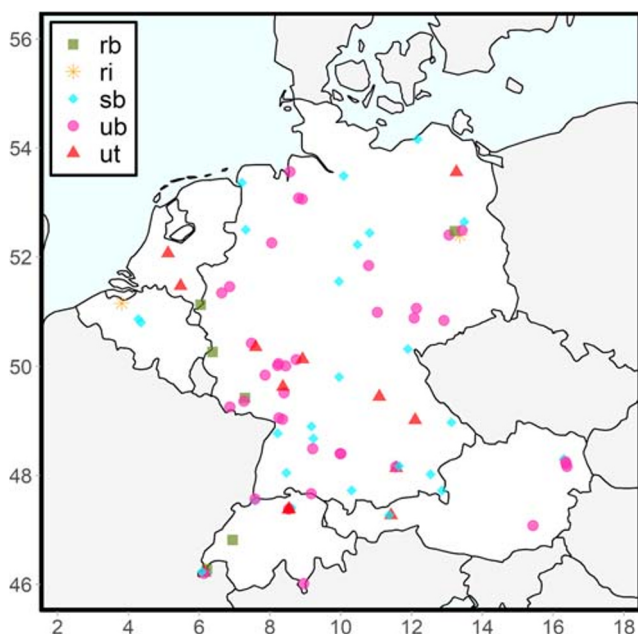

Fig. 1 Location of all 85 selected stations in Central Europe. Color and shape of points indicate the class of each station. Classes are specified as follows: rb = rural background, ri = rural industrial, sb = suburban background, ub = urban background, ut = urban traffic

Dee 2016), and earth system models (ESMs) as well as own analysis. The following large-scale meteorological variables formed the initial database: geopotential heights at the 850 hPa and 500 hPa levels (GH850, GH500), mean sea level pressure (MSLP), mean air temperatures at 850 hPa, 500 hPa, and 2 m (MT850, MT500, and MT2m), relative humidity at 850 hPa and 500 hPa (RH850, RH500), surface solar radiation downwards (SSRD), surface thermal radiation downwards (STRD), total cloud cover (TCC), and zonal and meridional winds at 850 hPa and 500 hPa (UWind850, UWind500, and VWind850, VWind500). Predictor data was retrieved with a $0.75° \times 0.75°$ resolution. Matching predictor data for projections under scenarios historical, RCP4.5, and RCP8.5 (CMIP5 label: historical, rcp45, rcp85) were extracted from seven CMIP5-ESMs listed in Table 1. RCP4.5 and RCP8.5 represent one intermediate and the highest representative concentration pathway with a radiative forcing stabilized at approximately 4.5 W/m^2 by or still rising, greater than 8.5 W/m^2 after 2100, respectively (for detail, refer to IPCC 2013).

The first run of the available ensemble was chosen for all variables. Thus, differences arising from the use of different ESMs were considered, whereas uncertainties from different initial conditions were not regarded. In accordance with the periods considered in the statistical downscaling models and projections, historical runs from 1993 to 2005 and scenario runs from 2006 to 2100 were used. Model data was regridded by nearest neighbor remapping to match the ERA5 resolution. Based on the geographic coordinates of each of the 85 stations, the mean of the nine grid boxes covering the area over and around the respective station location was calculated for all predictor variables.

The chosen statistical downscaling approach uses only observational data for both predictand and predictors (reanalysis) to train the statistical models. The models are later applied to ESM output assuming that the ESMs provide large-scale fields similar to the observed atmospheric variability. A simple linear scaling bias correction technique, used, e.g., by Gohar et al. (2017) and Teutschbein and Seibert (2012), was chosen and adapted to remove the monthly mean bias between reanalysis and climate model data.

The monthly mean difference between reanalysis and climate model data from 1993 to 2005 was used to bias-correct the respective ESM data from 1993 to 2100. Brands et al. (2013) assess the ESM performance to reproduce observed climatology not just on the basis of analyzing mean differences (bias), but also by examining distributional differences between reanalysis (ERA-Interim) and ESMs. Accordingly, to compare the distributional similarity between ERA5 reanalysis and ESM data, the two-sample Kolmogorov-Smirnov test (KS test) was applied. The used monthly time series were centered to have zero mean by subtracting the specific mean of the o-t-season from each timestep. Distribution differences were tested on the 95% significance level. To ensure a high consistency between ERA5 and ESM predictor data, all variables for which significant distributional differences remained for at least one station across all respective ERA5-ESM pairs were rejected. As a result, RH850, TCC, and UWind500 as well as UWind850 were hereafter neglected.

Indicated by previous research work, persisting pollution levels represent one of the most important drivers of ozone concentrations for specific locations in (Hertig et al. 2019) and across Europe (Otero et al. 2016). Thus, to account for prevailing pollution episodes, six MDA8O3 persistence metrics were added to the initial model input predictor set of each station. Persistence metrics were calculated by averaging the MDA8O3 concentrations of one to up to six preceding days.

Methods

Statistical downscaling models

Main drivers of o-t-events from April to September were analyzed by station-based statistical models representing the final results of two multistep modeling approaches. Initially,

Table 1 CMIP5 Earth System Models considered in this study

ESM	Institution	Atmospheric grid resolution
ACCESS1-0	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia	$1.25° \times 1.88°$
CanESM2	Canadian Centre for Climate Modelling and Analysis, Canada	$2.79° \times 2.80°$
CMCC-CMS	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy	$3.71° \times 3.75°$
CNRM-CM5	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France	$1.40° \times 1.41°$
IPSL-CM5-MR	Institut Pierre-Simon Laplace, France	$1.27° \times 2.50°$
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan	$2.79° \times 2.81°$
MPI-ESM-MR	Max Planck Institute for Meteorology, Germany	$1.87° \times 1.88°$

logistic regression (LR) was used to directly assess the probability of combined o-t-events. For comparison, multiple linear regression (MLR) was applied to assess at first separately the drivers of MDA8O3 concentrations and of TX values (henceforth named "MDA8O3-MLR" and "TX-MLR", respectively). Subsequently, the occurrence of o-t-events was deduced based on a combination of each single modeling result. According to the holistic European focus, a final European-wide predictor set was defined within both modeling processes. This set was then used across all stations to generate final station-based downscaling models. The importance of a predictor at each location is primarily expressed by its respective regression coefficient. Projections of o-t-events until the end of the twenty-first century were generated for each specific station based on these statistical downscaling models. To analyze health-relevant o-t-events under current but also future climatic conditions, the final predictor set should desirable meet the following criteria:

- contain at least one circulation dynamic, thermal, radiation, and humidity-based predictor as well as an ozone persistence metric to cover each climate change-relevant information content
- include only those predictors which carry physically meaningful and substantially relevant information of the predictand
- be based on unique, out-standing predictors; i.e., predictors of almost same information contents are eliminated to avoid overfitting the final station models

Standardized time series of all predictors entered the statistical models. Standardization based on respective means and standard deviations was done to ease interpretability of regression coefficients and thus the identification of main drivers. In order to achieve consistent and comparable results from the downscaling process, identical predictors were used for all predictands and modeling approaches, with the exception that no pollution persistence metrics entered the TX-MLR modeling process. Regression was performed within the free software environment for statistical computing R.

Logistic regression

Logistic regression (Peng et al. 2010; Wilks 2006) representing a common technique for probability forecasts was used to model the likelihood of o-t-events. For this approach, o-t-events were coded as binary time series with values of 0 (non-event) and 1 (event). LR is based on a logit transformation with the response not having to follow a normal distribution. To analyze the relationship between all predictors and the predictand variable and to identify the main drivers of o-t-events, lasso (lasso, least absolute shrinkage and selection) regression (Hastie et al. 2009; Hastie et al. 2015;

Tibshirani 1996; Tibshirani 2011) was used and embedded in a tenfold cross-validation process. Lasso regression tries to enhance prediction accuracy by shrinking or setting some coefficients to 0. Thus, lasso performs a variable selection leading to a subset of predictors showing the strongest impact on the predictand, accounting for multicollinearity of predictors. Hence, it is favored to ease interpretability of the statistical models. Within lasso, a tenfold cross-validation was used to determine tuning parameter t. The parameter t controls the amount of shrinkage applied to the estimates. Due to the highly imbalanced data sets, resulting from low numbers of exceedance days with mean and median non-event to event ratios of 5.3 (min 4.5, max 8.0) and 5.2, a stratified resampling method was embedded in the modeling process. Applying the SMOTE algorithm (Chawla et al. 2002) for unbalanced classification problems, an equally distributed dataset was generated (using R's DMwR package). The package glmnet was used for lasso regression. Predictors chosen in at least 90% of all cross-validation training models entered the final predictor set to build a lasso regression model per station based on all observations. As a result of the screening process, an optimized predictor set for every station was extracted. A final, unified European-wide predictor set was created by choosing only predictors selected by more than half of the station-specific LR models. Since predictors which carry the same information, but occur on different pressure levels (e.g., VWind500 and VWind850), entered the model building process, the variable with the higher mean absolute standardized regression coefficient across all models was finally chosen to limit similar information contents. Final station-based LR models were built based on the generated uniform, for whole Europe standardized final predictor set and later used for projections.

Multiple linear regression

MLR is considered as an effective tool to study the impact of predictors on the mean of the response variable. Thus, MDA8O3 concentrations and TX values were predicted separately and the results subsequently combined to assess the occurrence of o-t-events. The performance of this approach to model o-t-events was evaluated and compared to the LR model results. In accordance to the previously described approach, a similar multistep predictor selection process was used and adapted within the MLR model building process. The selection of predictors to create a specific optimum for each station was conducted through a backward regression procedure starting with models including all potential predictors. Step by step, the least important variables were sequentially removed from the regression equation according to the Akaike Information Criterion (AIC, Akaike 1974). To account for multicollinearity, variables with a variance inflation factor (VIF) larger than 10 were excluded from the equation

(James et al. 2013). The two separate final European-based predictor sets for the MDA8O3-MLR as well as TX-MLR models were created analogously to the LR approach.

Model performance

Typical evaluation metrics as accuracy and error rate are not suitable for highly imbalanced data sets when the primary goal is the identification of rare events (Branco et al. 2016). Performance of the LR statistical downscaling models was thus evaluated using various typical performance metrics for two-class situations and imbalanced datasets. The common and most often applied metrics especially in machine learning with imbalanced data sets are precision, recall and the F1-score (He and Ma 2013). Recall also named true positive rate (TPR) is used to measure the fraction of actual observed events that are correctly predicted, whereas the true negative rate (TNR) measured the fraction of actual observed non-events that are correctly predicted. Precision is used to measure the fraction of the predicted events that did actually occur. Precision and recall range from 0 to 1 with a perfect score of 1. The F1-score is based on precision and recall and gives equal weight to both measures. Besides these, various other metrics were used for model evaluation and are given in Table S3 in the Supplementary Material (Online Resource 2). For the MLR models, additionally, the coefficient of determination R^2 is used to determine the model performance. The performance evaluation of the final LR and MLR station models was also embedded in a tenfold cross-validation procedure.

Projections

Statistical projections of o-t-events and thus potential frequency shifts of these health-relevant occurrences until the end of the twenty-first century were assessed by comparing time slice differences under RCP4.5 and RCP8.5 scenario conditions. Therefore, 20-year periods in mid-century (2031–2050) and late-century (2081–2100) climate were compared to the base period (1993–2012). Twenty-year periods are suitable time slices because they are expected to be long enough to average out internal variability. Thus, air quality and temperature impacts representing the consequence of changes in forcing can be reliably assessed. Projections were conducted by exchanging the ERA5 predictor data used for modeling building with the corresponding ESM data. As a consequence of the carried out performance evaluation, the LR models were chosen for projections. More details on this will be given later in the "Statistical model performance and predictors" section. As ozone persistence predictor data with a daily resolution is not provided by the chosen ESMs, MDA8O3-MLR models were used to project the ozone pollution concentrations under future climate conditions. As previous day ozone concentrations represent a predictor in the MLR models themselves, a

proxy calculated by averaging all daily-based observations across all years in the base period for March 30th was used as a start value. The approach is accompanied by the fundamental assumption that future MDA8O3 concentrations in spring are similar to recently observed values. Several advantages exist by using an observation-forced proxy as start value. The accumulation of possible uncertainties and error sources based on the underlying model equations and simplifications are reduced by integrating an annual constant initial start value. Projecting the remaining MDA8O3 concentration levels in the o-t-season using ESM data also guarantees that the statistical relationships, their linked underlying processes, and causal relations regarding the predictors themselves as well as between predictors and predictand are accounted for. Projections of o-t-events could thus be interpreted against the background of stabilized ozone persistence conditions, with daily and monthly fluctuations presented in future ozone concentration predictor data. To evaluate the goodness of this approach, statistically modeled ozone values were used. Thus, MLR models based on only original predictor data and models including the start dummy are compared to each other. The results show a strong agreement across all models with a mean and median monthly correlation of 0.83 (min 0.35, max 0.98) and 0.86. Only seven stations (DEBB021, DEBE027, DEBE034, DEBE051, DEMV003, DEMV007, NL00639) showed significant distributional differences between both time series ($\alpha = 0.05$). Thus, MDA8O3 concentrations were projected by using the proxy as initial yearly start value across all 85 stations and seven ESMs as well as under both, RCP4.5 and RCP8.5, scenario conditions.

Results

Co-occurrence of health-relevant levels

The seasonal median value of TX was 21.07 °C with 25% and 75% of data being below 17.27 °C and 24.88 °C, respectively. All values were calculated as means over all 85 stations by only considering the months in the underlying o-t-season. According seasonal median MDA8O3 concentrations were 82.02 μg/m³ with 25% and 75% of data being below 65.68 μg/m³ and 101.34 μg/m³, respectively. For each station, mean TX and MDA8O3 values from April to September over the years are shown in Table S2 (Online Resource 1). The mean TX and MDA8O3 concentration level over Central Europe in the o-t-season was 20.97 °C and 84.85 μg/m³, respectively. Considering station characteristics, the upper 25% of MDA8O3 data exceeded on average the WHO AQG across all stations for all but two classes (ri and ut), highlighting the utility of the applied 75th percentile-based and 100 μg/m³ thresholds to define o-t-events.

The strong observation-based relationship between MDA8O3 concentrations and TX values in the o-t-season is shown in Fig. 2. Significant (α = 0.05), high monthly correlations from April to September become visible for all 85 stations. A high frequency of co-occurring values of MDA8O3 and TX exceeding the 75th thresholds in the base period became apparent. Elevated TX values were often accompanied by high MDA8O3 pollution levels; on average, 66% of days with elevated thermal load also showed MDA8O3 values surpassing the 75th threshold (min 50, max 74, median 68). Thus, thermal load occurrences were associated with coinciding O$_3$ pollution. Averaged across all 85 stations, on 592 days (min 380, max 671) in the base period (approx. 16% of all days in the o-t-season), combined 75th threshold exceedances of both target variables were observed with a median of 602 days. The spatial distribution over Central Europe, shown in Fig. 3, revealed a higher absolute and relative amount of these occasions especially in central to northern parts of the study region.

Taking the additional threshold of 100 µg/m^3 used to define combined o-t-events into account, the occurrence of thermal load was connected with co-occurring MDA8O3 exceedances on average on 75% of days (min 50, max 88, median 75). A total of 670 days (ca. 18% of all days in the o-t-season) with o-t-events (min 381, max 802, median 674) were overserved on average per station. As may be expected in accordance with the "Introduction" section presented analysis (e.g. EEA 2019) and the described ozone formation in the presence of solar radiation, the spatial distribution (not shown) revealed that especially stations in southern Central Europe show a higher proportion of heat days with MDA8O3 exceedances and a higher frequency of o-t-events. Thus, at these stations, MDA8O3 concentrations exceeded more often the WHO AQG.

Statistical model performance and predictors

The final predictor set of the LR models subsequently used for projections contained the following meteorological variables: GH850, MT850, RH500, STRD, and VWind500. Beside these meteorological variables, only ozone pollution levels of the first previous day of all initial chosen persistence metrics were identified as one main driver of o-t-events, henceforth named "LagO3". Thus, the final predictor set contained at least one circulation dynamic, thermal, radiation, and humidity-based predictor as well as an ozone persistence metric, all identified to have a strong effect on the probability of occurring o-t-events. Regarding statistical model performance, the mean and median F1-score of the LR model results across all stations was 0.63 (min 0.49, max 0.74). Thus, the models showed in general an acceptable performance to predict o-t-events. The TPR and TNR means and medians amounted to 0.83 (min 0.74, max 0.89) and 0.84 as well as 0.81 (min 0.75, max 0.85), respectively. Hence, a similar, balanced performance of the models was achieved to identify known events and non-events. Additional results concerning LR model performance analysis and evaluation can be found in Table S3 in the Supplementary Material (Online Resource 2). In comparison, the second model approach based on the combination of results of the two separate MLR models showed considerably less performance to model o-t-events. These final predictor sets for the MLR models sets were both composed of RH500, SSRD, and VWind500. For MDA8O3-MLR models, LagO3 was additionally selected within the predictor screening process. The cross-validated mean and median F1-score amounted to 0.36 (min 0.25, max 0.45) across all stations. As the respective MLR models aimed to assess the mean of the TX and MDA8O3 response, not surprisingly, on average, TNR values with a mean and median of 0.94 (min 0.85, max 0.97) were accompanied by low TPR values with a mean and median of 0.28 (min 0.17, max 0.37) and 0.27. As might be expected, the high TNR in comparison to the low TPR values indicated in general the inability of the models to assess elevated levels of both target variables. Since the LR models clearly outperformed the combined MLR models in capturing o-t-events, only the first model approach was used to analyze and project future
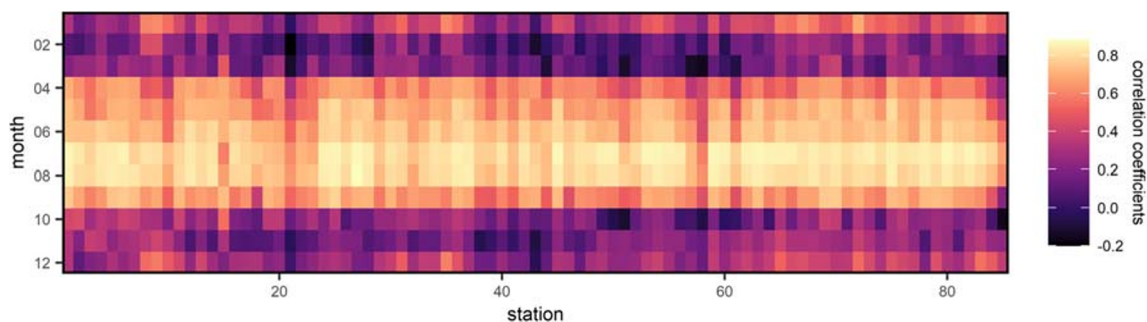

Fig. 2 Spearman rank correlation coefficients between observed MDA8O3 concentrations and TX values for each station in Central Europe during the 1993–2012 period. Station numbers depicted on the *x*-axis are in accordance with Table S2 in the Supplementary Material (Online Resource 1)
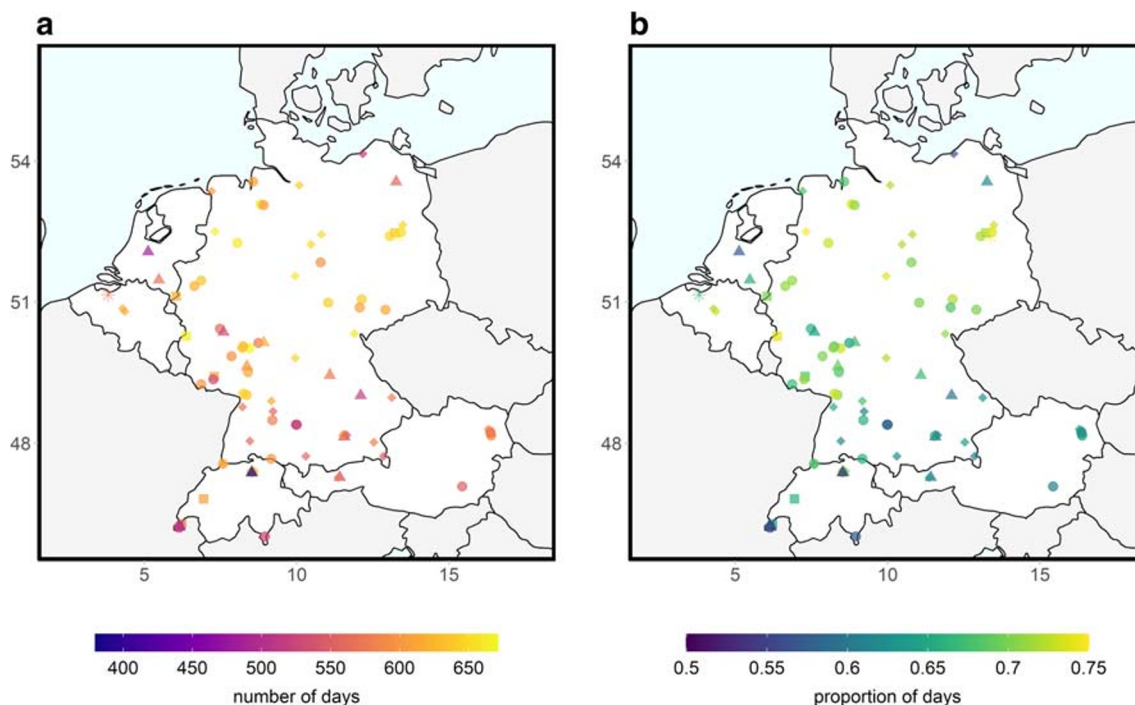
Fig. 3 Spatial distribution of the number of days with observed combined exceedances of MDA8O3 and TX values over the 75th threshold (**a**) and the proportion of these event days to the total number of days with thermal load (**b**)

frequency shifts in these combined events. As the MDA8O3-MLR models were applied to generate the LagO3 predictor for subsequent statistical projections, their ability to assess the mean of the surface daily maximum ozone response was additionally evaluated. The co-efficient of determination R^2 ranged between 0.43 and 0.73 with mean and median values of 0.58 across Central Europe. The spatial distribution of the F1-score for LR models and R^2 for MDA8O3-MLR models is shown in Fig. 4. Although model performances were satisfactory in general over Central Europe to reliably project future MDA8O3 concentrations and frequency shifts in o-
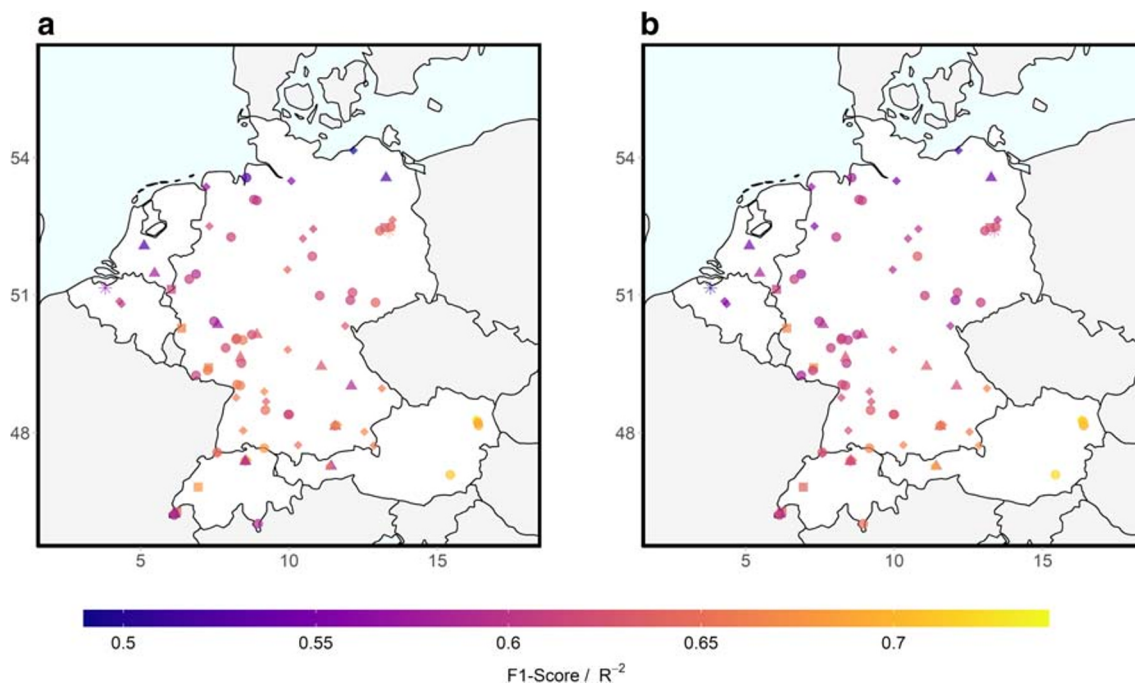

Fig. 4 Spatial distribution of the F1-score for LR (**a**) and of R^2 for MDA8O3-MLR (**b**)

t-events, station-specific findings should always be treated with caution and be evaluated against the background of the respective individual model quality.

As MDA8O3-MLR models are used for projecting MDA8O3 concentrations and LR models to assess future frequency shifts in the occurrence of o-t-events, the two model approaches build the focus of the following sections.

Main Drivers

Drivers of MDA8O3

The strength and kind of a relationship between a predictor and a respective target variable can be interpreted in terms of the magnitude and the sign of the predictor's standardized regression coefficient. Consequently, LagO3 was identified as the most important driver (MID) of MDA8O3 concentrations not just on average across Central Europe with a mean standardized coefficient of 17.23 (min 11.93, max 26.66, median 16.95), but also as the most important driver at each single station. The respective mean coefficient for SSRD was 4.37 (min 1.74, max 10.88, median 3.91) representing the second most important driver ("SMID") across Central Europe. The predictors VWind500 and RH with values of − 2.43 [min − 8.55, max 1.70, median − 2.65) and − 1.79 (min − 3.16, max 1.09, median − 1.92) showed in general a smaller and negative influence. Thus, for example, rising relative humidity levels reduce in general MDA8O3 concentrations. Since tropospheric ozone is formed in the presence of sunlight based on the reaction of precursor gases showing an increased reactivity by rising air temperatures and solar radiation, the statistical model equations capture the underlying physical processes of O_3 formation. As air temperature can mostly be regarded as a proxy for solar radiation in the o-t-season, the identification of SSRD as one of the most important drivers within the study region highlights and confirms previous findings identifying the strong relationship between temperature and ozone (Coates et al. 2016; Jacob and Winner 2009; Oswald et al. 2015). The overall found impact of VWind500 is in good agreement with earlier studies showing that wind speed and direction influence ozone concentrations (Dueñas et al. 2002; Gardner and Dorling 2000; Husar and Renard 1998). Elevated relative humidity values indicate increased instability and cloudiness and thus reduced incoming solar radiation. The negative impact of RH on O_3 concentrations was also specified in previous studies (Camalier et al. 2007; Demuzere et al. 2009; Dueñas et al. 2002). The spatial distribution of the SMID and TMID reflect the importance of different MDA8O3 governing mechanisms. In the western parts of Europe, VWind500 plays a major role, indicating that inflow from remote sources is important for MDA8O3 concentrations, while in the eastern parts of Europe, in situ radiation-related predictors are primarily related with spring and summertime ozone concentrations.

The identification of main drivers in general concurs well with Otero et al. (2016). By examining the influence of synoptic and local meteorological conditions on MDA8O3 concentrations over Europe, the authors found evidence that next to ozone persistence also some meteorological parameters play an important role in explaining most of the observed ozone variance, e.g., maximum temperature, relative humidity, and solar radiation. Figure 5 shows the spatial distribution of the identified second and third most important driver ("TMID") of MDA8O3 concentrations in Central Europe.

Drivers of combined events

Table 2 gives an overview of the averaged standardized regression coefficients over Central Europe for the LR models. The results emphasize the physical validity of the models as MT850 and LagO3 were in general identified as the strongest influencing factors of o-t-events. The spatial distribution of main drivers in Central Europe, grouped by the most (a), second most (b), and third most important drivers (c), is shown in Fig. 6. The respective odds ratio (OR) is depicted for each driver (d–f) representing the odds that an o-t-event will occur given the presence of the specific main driver, compared to the odds of the event occurring in the absence of that predictor. Thus, OR above 1 indicate higher odds of an o-t-event to appear. Respectively, OR below one lower the odds of an o-t-event.

It became apparent that LagO3 was the MID in northwestern Central Europe. Hence, the persistence of ozone pollution, i.e., prevailing high pollution levels due to an incomplete degradation of previous day concentrations of O_3, may play a substantial role in driving MDA8O3 concentrations in these regions. Persistence can cause long-lasting or day-to-day increases of high concentration levels leading to persistently elevated ozone levels on subsequent days. MT850 represented the MID for the majority of stations in southern to eastern Central Europe with some predictors showing occasional elevated to extremely high OR (d). At these stations, it became evident that o-t-events could mainly be explained by mean temperature levels above the boundary layer at the 850 hPa level.

STRD was in general the third most important impactor factor (Table 2) in Central Europe, with low radiation values favoring the occurrence of o-t-events. This is also supported by the low OR < 1 (Fig. 6e and f) shown for stations selecting STRD as the second and third most important driver. STRD is mainly controlled by water vapor and aerosols such as cloud water droplets as well as mainly determined by the shallow layer close to the surface. Hence, it may be a proxy for high specific, thermal radiation-relevant trace element loadings and
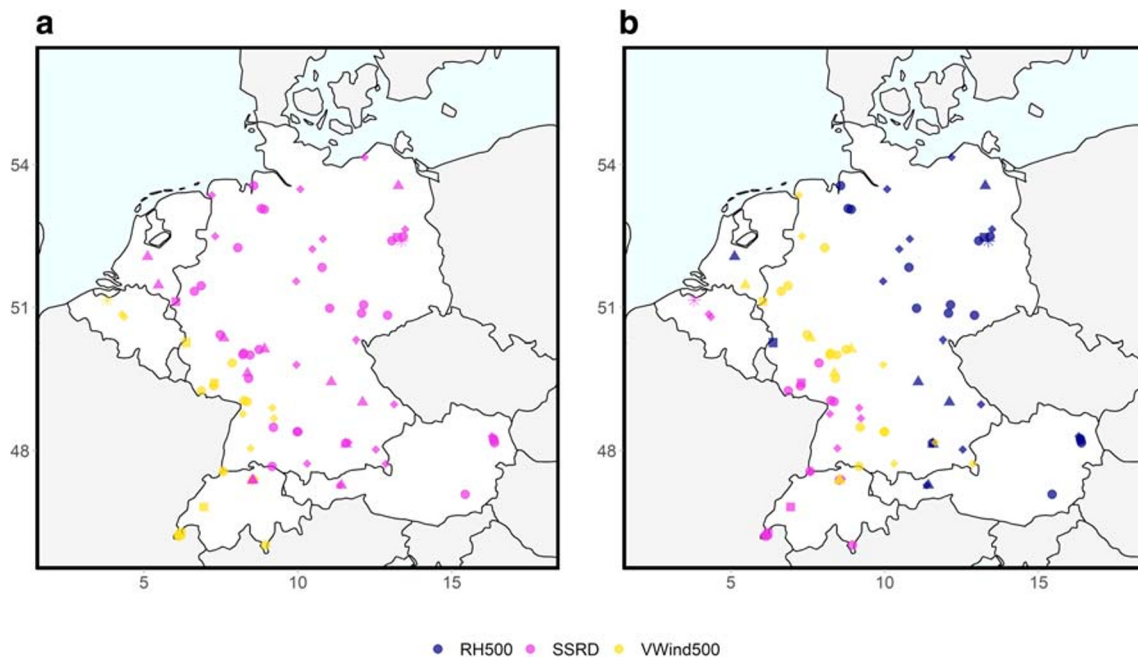
Fig. 5 Spatial distribution of the second most important (**a**) and third most important driver (**b**) of MDA8O3 in MLR. The shape of the legend symbols has no meaning

pollution events as well as moist and humid conditions. Thus, days with lower STRD favored the occurrence of o-t-events.

VWind500 is an indicator of airflow conditions, with positive values representing wind from southern directions. GH850 represents a circulation dynamic predictor with low values also being considered as indicator for cold, moist, and humid conditions. VWind500 was only selected twice as TMID and GH850 were occasionally selected as SMID and TMID at stations in the northwestern part of Central Europe with, according to the OR, low wind levels, and high geopotential heights (associated with low humidity values) favoring the occurrence of o-t-events. These findings are not just in good agreement with the results of the MDA8O3-MLR models at these stations, but also confirm the underlying physical processes and key factors known to determine temperature and ozone levels. The relevance of these drivers at the specific stations may also be associated with their location in

the westerly wind zone close to the North Atlantic. Transport of moist air masses from the ocean inlands leads generally to increased relative humidity levels, the evolution of clouds, and precipitation in coastal regions. No distinct differences were found regarding station characteristics, so main drivers of o-t-events are primary related to synoptic influences rather than individual station classes.

Projections for the twenty-first century

Projections of MDA8O3

A comparison of values statistically modeled using historical ESM predictor data with reanalysis-modeled data in the historical period from 1993 to 2005 showed good agreement. Monthly biases were extracted by calculating the monthly mean differences between the statistically modeled values using either historical ESM predictor data or reanalysis-based data. Biases were normalized by the standard deviations of the reanalysis-based data. Throughout this section, numbers in brackets refer to minimum and maximum values grounded on ESM projection results averaged across all stations. Mean and median monthly biases across all ESMs amounted to 0.28 (min − 0.60, max 0.66) and 0.24, respectively. Across all seven ESMs, only one station (AT900ZA) showed for two ESMs significant monthly distributional differences between the statistically modeled time series ($\alpha = 0.05$).

As the downscaling models per se showed a good performance in the historical period over Central Europe, statistical MDA8O3 projections were conducted by exchanging the

Table 2 Standardized regression coefficients in LR for each predictor variable. Numbers are the mean and median across all stations; minimum and maximum values are shown in brackets

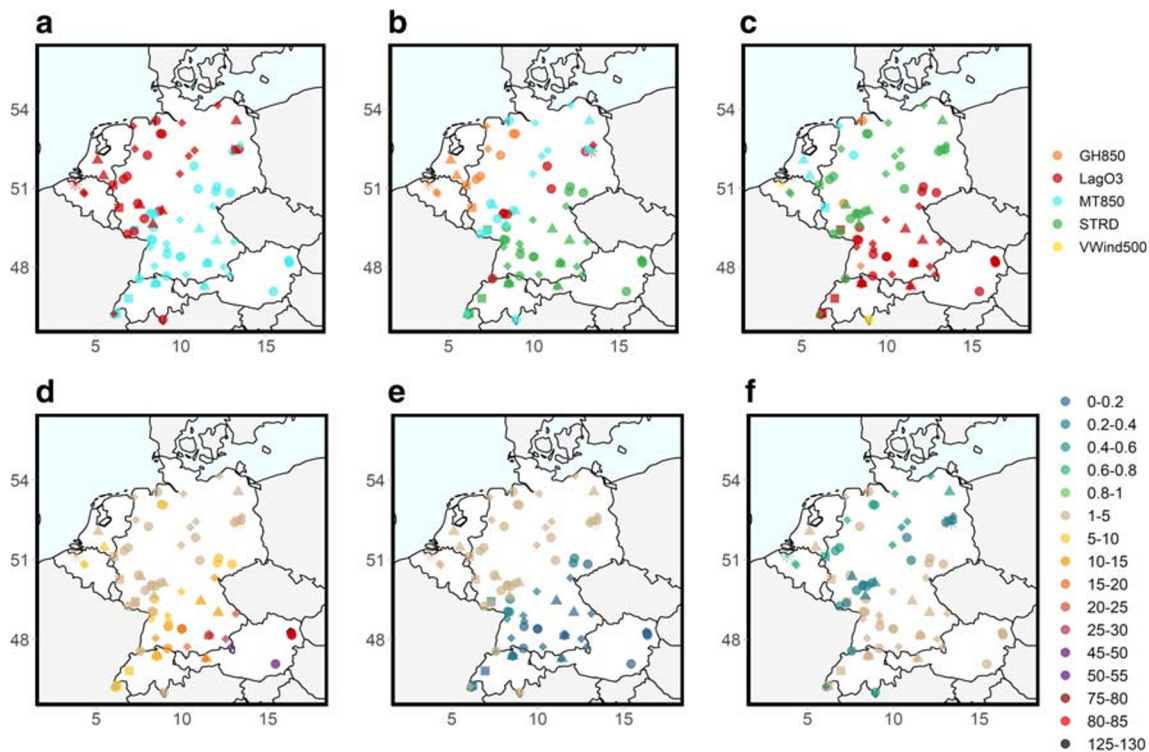| Predictor	Mean	Median
GH850	0.64 (0.13, 1.15)	0.69
LagO3	1.21 (0.82, 1.68)	1.22
MT850	1.78 (0.12, 4.86)	1.57
RH500	− 0.24 (− 0.65, 0.15)	− 0.24
STRD	− 1.30 (− 2.91, −0.30)	− 1.24
VWind500	− 0.34 (− 0.98, 0.43)	− 0.35

Fig. 6 Spatial distribution of the most important (**a**), second most important (**b**), and third most important (**c**) driver of o-t-events in LR. The odds ratio is shown for the respective identified main driver at each station (**d–f**). The shape of the legend symbols has no meaning

ERA5 predictor data with corresponding ESM data. MDA8O3 values were modeled for 1993 to 2100 under both scenario conditions. Considering the ensemble mean over Central Europe across all seven ESMs under RCP4.5 scenario assumptions, a mean and median change of − 0.19% (min − 9.84, max 35.12) and − 0.59% for the mid-century as well as − 0.10% (min − 11.68, max 40.15) and − 0.54% for the late-century period were assessed. Thus, on average, no relevant overall change became apparent, under the assumption of stationarity of the start value of MDA8O3 concentrations as well as of governing statistical relationships. Similar results became apparent for RCP8.5 scenario conditions with a mean and median change of 1.09% (min − 18.65, max 19.74) and 1.05% for the mid-century as well as − 0.32% (min − 25.27, max 12.68) and − 0.22% for the late-century period. With respect to spatial distributions, no distinct pattern became evident with only one station (CH0011A) showing elevated multi-model mean increases of MDA8O3 concentrations of 24.54% (mid-century) and 27.34% (late-century) under RCP4.5 and multi-model mean decreases of 0.01% (mid-century) and 16.06% (late-century) under RCP8.5 scenario assumptions. Although the projections for this station were comparably strong across all applied ESMs, the very different ensemble-mean percentage changes of this station for almost all time slices were highly determined by the results of one ESM (IPSL-CM5A-MR) with a modeled change of over 185% and 210% (RCP4.5) and over − 70% and − 80%

(RCP8.5), respectively. As the multi-model mean of this station is highly dependent on one model outlier, a note of caution should be sounded with regard to projections for this station. In summary, stabilized MDA8O3 concentrations throughout the twenty-first century were assessed for Central Europe under both scenario conditions and the previously outlined, underlying modeling assumptions showing no relevant spatial differences across the vast majority of stations.

Projections of combined events

Table 3 provides an overview of the multi-model time slice differences, statistically downscaled under RCP4.5

Table 3 Changes (%) regarding the amount of days with o-t-events between the periods 2031–2050 and 2081–2100 compared with 1993–2012 under RCP4.5 and RCP8.5 scenarios. Numbers refer to the ensemble mean and median across all seven ESMs. Numbers in brackets give minimum and maximum changes based on single ESM projections averaged across all 85 stations

Scenario	Period	Mean (%)	Median (%)
RCP4.5	2031–2050	8.94 (− 0.39, 22.70)	7.80
	2081–2100	16.84 (− 1.23, 39.77)	16.22
RCP8.5	2031–2050	13.33 (0.04, 33.18)	11.37
	2081–2100	37.52 (− 1.85, 96.42)	33.92

and RCP8.5 scenario assumptions. It is evident that projections point to an on average growing number of days with concurrent occurrences of thermal load and ozone pollution in Central Europe. Not surprisingly, considering the identified main drivers in LR and the stronger radiative forcing by greenhouse gases and the related enhanced rise in global mean temperatures, comparably stronger frequency shifts were assessed under RCP8.5 scenario assumptions for mid- and late-century climate conditions. For example, a substantial projected increase of o-t-events in the period 2081–2100 compared to the period 1993–2012 was revealed for both scenarios, but for RCP4.5 with about 17% only being less than half as strong as for RCP8.5 with about 38%.

The shown multi-model results were based on the climate change signal of seven individual ESMs showing for almost all periods the same sign, but a different magnitude of change. The respective projected changes of each single ESM can be found in Table S4 in the Supplementary Material (Online Resource 2).

An overview of the spatial distribution is provided in Fig. 7 showing the multi-model mean changes for every single station under RCP4.5 (upper) and RCP8.5 (lower) scenario conditions for mid- (left) and late-century (right section of figure) climate. In general, stronger frequency shifts over Central Europe for late-century and under RCP8.5 scenario conditions became apparent. The single most striking observation to emerge from the spatial analysis was the clear identification of hotspot regions across all scenarios and periods. In general, the respective strongest increases in o-t-events were projected for south- to mid-eastern Central Europe. Under RCP4.5 scenario, assumptions projected changes ranged from − 0.39 (BETR740) to 22.70% (DEBY079), with 83 stations for mid-century climate showing an increase of o-t-event days. For comparison, late-century changes ranged from − 1.23 (BETR740) to 39.77% (DEBY088) with a growing number of event days projected for 84 stations. Similar results became apparent under RCP8.5 scenario conditions. For the mid-century period, all 85 stations showed an increase in o-t-events. Projected changes reached from 0.04 (DEHB002) up to 33.18% (DEBY079). Until the end of the twenty-first century, a higher frequency of o-t-events was assessed for 84 stations, ranging in general between − 1.85 (BETR740) and 96.42% (DEBY063).

The stations showing a late-century increase of at least 60% (16 stations) under RCP8.5 were all located in the German federal states of Bavaria and Baden-Wuerttemberg or close to the southern German border in Austria and Switzerland. As in general south- to central-eastern parts of the study region could be identified as hotspots of regional change, a low climate change signal became apparent in general at stations in the northern and northwestern part of Central Europe. As shown in Fig. 6, the most influencing drivers for the vast majority of stations in the identified hotspot regions are MT850, LagO3, and STRD according to the LR models. The MDA8O3-MLR models assessed stabilized MDA8O3 concentrations throughout the twenty-first century (refer to chapters Projections and Projections of MDA8O3 for detail). These values entered later as LagO3 predictor the LR model-based projection process. Consequently, primary projected changes of MT850 and STRD may mainly affect the projected frequency increases of o-t-events under both scenario assumptions, also highlighted by the substantial OR. In agreement with the previous analysis of main drivers of o-t-events, other than regional hotspots, no significant and consistent dependence of the changes from station characteristics became apparent. Additional analysis about monthly mid-century to late-century percentage changes in the future occurrences of o-t-events based on multi-model means under RCP4.5 and RCP8.5 can be found in the Supplementary Material (Figures S1–S4, Online Resource 3). In summary, the downscaling results indicated a considerable increase in the occurrence of combined o-t-events from April to September and hence of thermal and air pollution load in the future. Regional hotspots of rising frequencies of o-t-events became apparent. Projected high frequency shifts became primarily evident in south- to mid-eastern regions of Central Europe.

Discussion

The occurrence of combined ozone-temperature-events for Central Europe based on the selection of 85 stations with different air quality settings in Austria, Germany, Belgium, The Netherlands, and Switzerland was assessed. The primary presumption of a strong positive correlation between surface daily maximum temperature and surface daily maximum ozone concentrations was confirmed for all analyzed stations. Due to the photochemical ozone formation with increased reactivity at higher temperatures and enhanced solar radiation, this relationship was evident primarily within the defined o-t-season from April to September. Burden-inducing, combined were defined by quantile (75th) as well as threshold exceedances (100 μg/m^3). Different modeling approaches were tested to model the occurrence of combined o-t-events. As a result of an extensive model performance evaluation, statistical downscaling models based on logistic regression were the final chosen modeling approach in this context. The models were at first developed to assess the relationship between large-scale predictors, ozone pollution levels, and local-scale events in the observational period. A holistic approach for Central Europe was chosen, accompanied by the greatest possible standardization of the model building process across all stations. Using a screening procedure primary based on lasso regression, a unified predictor set was assessed for all individual LR station models over Central Europe.
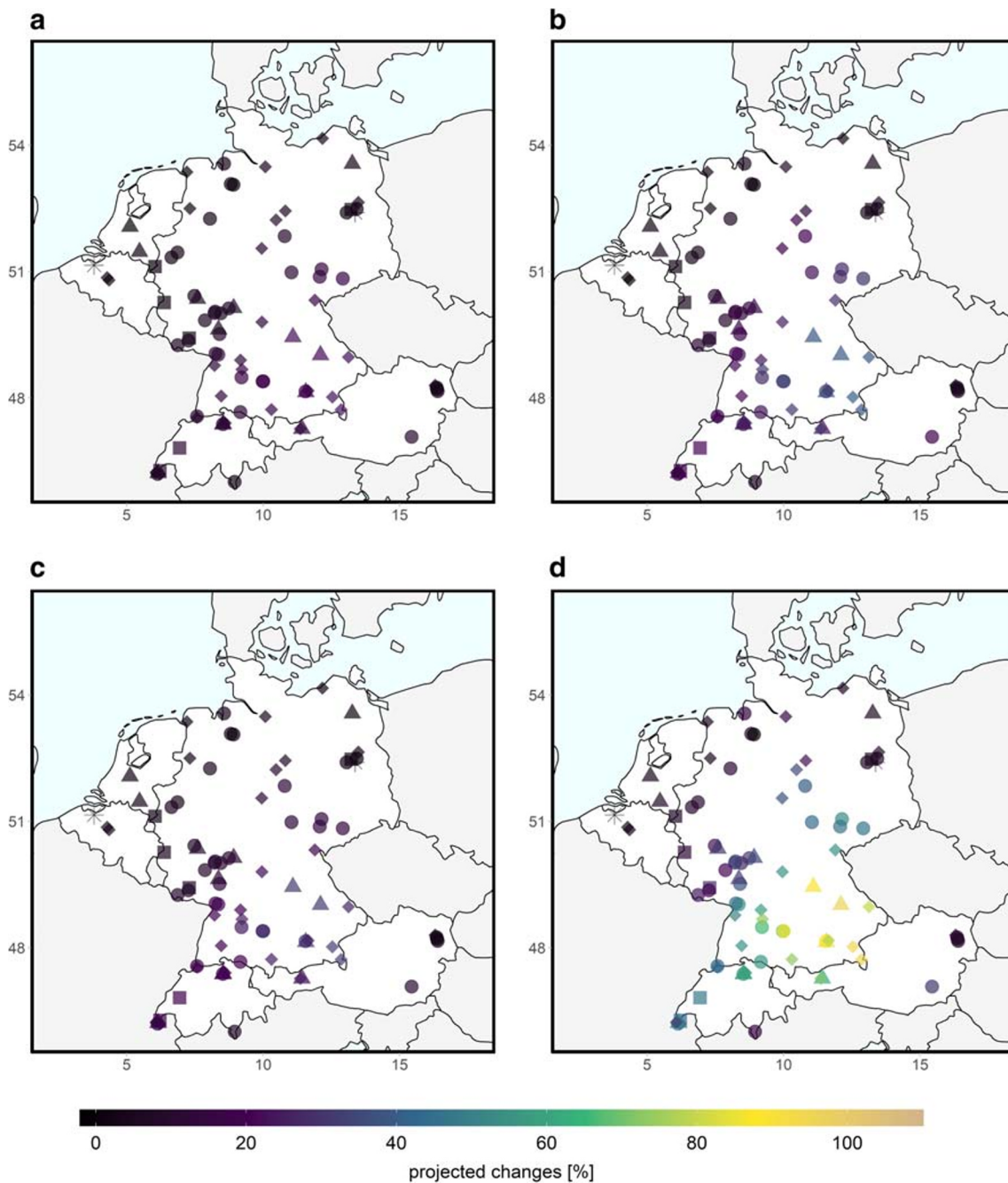
Fig. 7 Spatial distribution of projected changes (%) with respect to the number of days with o-t-events between the periods 2031–2050 (left) and 2081–2100 (right) compared with 1993–2012 under RCP4.5 (top) and RCP8.5 (bottom) scenario. Numbers refer to the ensemble mean across all seven ESMs

MT500, STRD, and LagO3 were found to be the main drivers of co-occurring high levels of MDA8O3 concentrations and TX levels.

The LR models were then used for projections under future scenarios. For this purpose, ESM data entered the statistical models and projections of o-t-events under RCP4.5 and RCP8.5 scenario assumptions to account for the ongoing climate change until the end of the twenty-first century. As the seven chosen ESMs of the CMIP5 project did not provide

surface daily maximum ozone levels, a MLR modeling approach to generate MDA8O3 projections until the end of the twenty-first century subsequently integrated as LagO3 predictor in the LR models was used. Multi-model mean changes (%) in the occurrence of o-t-events of 8.94% and 16.84% for mid- and late-century climate in comparison to 1993–2012 were found for the RCP4.5 scenario across all stations. Accordingly, changes of 13.33% and 37.52% under RCP8.5 scenario assumptions were generated. Regarding spatial

distributions, regional hotspots in south- to central-eastern parts of Central Europe became apparent, but no obvious dependence of the projections on station classes was evident.

Concerning this study, it is plausible that a number of limitations may have influenced the results obtained. All findings need to be interpreted with regard to the chosen holistic focus of the study on Central Europe as the main spatial scale. In order to identify more subtleties and differences within the study area and thus to highlight, e.g., station class-specific differences, a more individual approach would have to be used. An initial focus of the study was to generate a homogenized database based on preselected ozone and temperature station pairs with similar boundary conditions (e.g., altitude levels) showing a strong relationship between ozone and temperature. To incorporate a sufficient number of stations spread across the study region, no further selection process based on the types of air quality settings (e.g., urban traffic stations vs. rural background stations) was incorporated. Consequently, the direct influence of emissions of precursors and hence reactive pollutants affecting ozone production and depletion as well as air pollution transport processes on varying spatial scales were not further considered and incorporated in the analysis. A more detailed and differentiated approach concerning the varying ozone production chemistry and underlying processes should be integrated in further analysis. As the investigations presented here have mainly focused on an European-wide scale, for comparison, a more regional- and station-specific approach should be the target of future work. Thus, along with a more individualized predictor screening process, final station models based on their station-specific predictor optimum could be generated and applied for projections. As a result, frequency shifts of o-t-events until the end of the twenty-first century based on station-specific models and on the here presented unified approach could be compared and relevant differences assessed. Additionally, combined events for both underlying target variables could be more station- or region-specific and might not just integrate worldwide air quality guidelines and percentile thresholds but also use, e.g., individual recommendations of national- or regional-wide institutions. This would also account for current and future demographic and health-relevant developments of the Central European population, as, e.g., aging and pre-existing conditions lead to a more vulnerable population to thermal load and concurrent air pollution.

A major source of uncertainty is based on the method applied to assess future MDA8O3 concentrations later used for projecting o-t-events. Even though the statistically modeled values were in good agreement with observations and reanalysis-modeled values, the assumed stationarity of the start value of MDA8O3 concentrations as well as the performance of the MLR models themselves influenced and may lead to inaccuracies in the subsequent projections. The performance of the generated MDA8O3-MLR models was in

general satisfactory for Central Europe, but slightly disappointing for some individual stations. This is to a large part due to the objective to find holistic downscaling models with a fixed predictor set for all stations, but it has to be especially kept in mind when interpreting station-specific developments and projections. The here presented findings should be read with regard to the chosen framework conditions of the study. The projected MDA8O3 levels should be interpreted to be a result of recent and future mitigation strategies and policies not comprising new or other sources of precursor emissions counteracting or strengthening these measures under future European climate conditions. Hence, the results must be interpreted against the background of these stabilized projected MDA8O3 concentrations entering as LagO3 predictor the LR models to project future o-t-events under both scenario assumptions. Consequently, as comparably stable ozone persistence conditions are assumed, changes of all meteorological predictors—mostly future projected MT850 and STRD conditions showing the highest standardized regression coefficients in the LR models (see Table 2)—may strongly affect the general projected increase in future o-t-events across Central Europe.

One negative factor regarding the projections for the twenty-first century is inevitable and is connected to the in general coarse resolution of the seven ESMs being regridded to match the spatial resolution of the ERA5 reanalysis data. Data errors and inaccuracies might have emerged due to interpolation, processing, generalizations, and uncertainties in the CMIP5 climate projections themselves. Associated limitations were accounted for by not just only bias-correct the data but as well as pre-select predictor data based on significant distributional differences between reanalysis and ESM data. Even if the initial shortcomings and limitations of the ESM data are accounted for in the predictor preparation, modeling, and projection process, the use of ESMs with higher spatial resolution would be beneficial. Thus, future work needs to be undertaken by means of global climate model simulations of subsequent Coupled Model Intercomparison Project phases (i.e., CMIP6). Furthermore, longer MDA8O3 and TX time series with more recent observational data should be integrated in future studies to confirm the modeled relationships and to assess even more reliable statistical downscaling models and projections.

Conclusion

In general, a strong correlation between daily maximum ozone concentrations and temperature values was found from April to September in Central Europe accompanied by a high frequency of concurrent elevated levels of both variables. MLR and LR models describing the relationship between large-scale meteorological predictors, prevailing high pollution

levels and local-scale events, performed well enough to become practical tools for predicting daily MDA8O3 concentrations as well as combined ozone-temperature events. Projections of o-t-events assuming stationary of the statistical relationships under historical and scenario conditions until the end of the twenty-first century were evaluated for the 2031–2050 and 2081–2100 time windows. A sharp increase of o-t-events was projected under RCP4.5 and RCP8.5 scenario assumptions. It became evident that south- to central-eastern parts of the study area represented hotspot regions with more frequent occurrences of these combined events in the o-t-season. High levels of ozone and temperature will increasingly coincide in these areas, thus posing an even intensified threat to human life as a result of their associated individual and combined health effects. Special attention should be paid to these vulnerable regions by the formulation and implementation of consensual European climate change mitigation strategies. MT500, LagO3, and STRD can be considered as powerful predictors to assess days with concurrent thermal and air pollution load and should consequently be interpreted as main drivers of o-t-events. GH850, VWind500, and RH500 became also apparent to substantially influence co-occurring elevated levels of surface daily ozone and temperature levels. The results suggest that ozone persistence is particularly relevant for subsequent pollution levels.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Code availability Not applicable

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19:716–723. https://doi.org/10.1109/TAC.1974.1100705

Analitis A, Michelozzi P, D'Ippoliti D, De'Donato F, Menne B, Matthies F, Atkinson RW, Iñiguez C, Basagaña X, Schneider A (2014) Effects of heat waves on mortality: effect modification and confounding by air pollutants. Epidemiology 25:15–22. https://doi.org/10.1097/EDE.0b013e31828ac01b

Anderson GB, Bell ML (2011) Heat waves in the United States: mortality risk during heat waves and effect modification by heat wave characteristics in 43 US communities. Environ Health Perspect 119:210–218. https://doi.org/10.1289/ehp.1002313

Baccini M, Biggeri A, Accetta G, Kosatsky T, Katsouyanni K, Analitis A, Anderson HR, Bisanti L, D'Ippoliti D, Danova J (2008) Heat effects on mortality in 15 European cities. Epidemiology 19:711–719. https://doi.org/10.1097/EDE.0b013e318176bfcd

Black E, Blackburn M, Harrison G, Hoskins B, Methven J (2004) Factors contributing to the summer 2003 European heatwave. Weather 59:217–223. https://doi.org/10.1256/wea.74.04

Bloomer BJ, Stehr JW, Piety CA, Salawitch RJ, Dickerson RR (2009) Observed relationships of ozone air pollution with temperature and emissions. Geophys Res Lett 36:L09803. https://doi.org/10.1029/2009GL037308

Branco P, Torgo L, Ribeiro RP (2016) A survey of predictive modeling under imbal-anced distributions. ACM Comput Surv 49:1–48. https://doi.org/10.1145/2907070

Brands S, Herrera S, Fernández J, Gutiérrez JM (2013) How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa? Clim Dyn 41:803–817. https://doi.org/10.1007/s00382-013-1742-8

Camalier L, Cox W, Dolwick P (2007) The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. Atmos Environ 41:7127–7137. https://doi.org/10.1016/j.atmosenv.2007.04.061

Carro-Calvo L, Ordóñez C, García-Herrera R, Schnell JL (2017) Spatial clustering and meteorological drivers of summer ozone in Europe. Atmospheric Environment 167:496–510. https://doi.org/10.1016/j.atmosenv.2017.08.050

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357. https://doi.org/10.1613/jair.953

Coates J, Mar KA, Ojha N, Butler TM (2016) The influence of temperature on ozone production under varying NOx conditions–a modelling study. Atmos Chem Phys 16:11601–11615. https://doi.org/10.5194/acp-16-11601-2016

Colette A, Andersson C, Baklanov A, Bessagnet B, Brandt J, Christensen JH, Doherty R, Engardt M, Geels C, Giannakopoulos C, Hedegaard GB, Katragkou E, Langner J, Lei H, Manders A, Melas D, Meleux F, Rouïl L, Sofiev M, Soares J, Stevenson DS, Tombrou-Tzella M, Varotsos KV, Young P (2015) Is the ozone climate penalty robust in Europe? Environ Res Lett 10:084015. https://doi.org/10.1088/1748-9326/10/8/084015

Demuzere M, Trigo RM, Vila-Guerau de Arellano J, Van Lipzig N (2009) The impact of weather and atmospheric circulation on O3 and PM10 levels at a rural mid-latitude site. Atmos Chem Phys 9:2695–2714. https://doi.org/10.5194/acp-9-2695-2009

Deutscher Wetterdienst (DWD) (2020) Wetterlexikon. Examples can be retrieved from: https://www.dwd.de/DE/service/lexikon/lexikon_node.html. Accessed 25 Apr 2020

Dueñas C, Fernández MC, Cañete S, Carretero J, Liger E (2002) Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean Coast. Sci Total Environ 299:97–113. https://doi.org/10.1016/S0048-9697(02)00251-6

Eis D, Helm D, Laußmann D, Stark K (2010) Klimawandel und Gesundheit-Ein Sachstandsbericht. Robert Koch-Institut, Berlin

European Environment Agency (EEA) (2014) AirBase - the European air quality database. Data and metadata available at: https://www.eea.europa.eu/dataand-maps/data/airbase-the-european-air-quality-database-8. Accessed 26 Mar 2020

European Environment Agency (EEA) (2019) Air quality in Europe — 2019 report. European Environment Agency, Luxembourg

Fang Y, Mauzerall DL, Liu J, Fiore AM, Horowitz LW (2013) Impacts of 21st century climate change on global air pollution-related premature mortality. Clim Change 121:239–253. https://doi.org/10.1007/s10584-013-0847-8

Fiore AM, Naik V, Leibensperger EM (2015) Air quality and climate connections. J Air Waste Manage Assoc 65:645–685. https://doi.org/10.1080/10962247.2015.1040526

Gardner MW, Dorling SR (2000) Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmos Environ 34:21–34. https://doi.org/10.1016/S1352-2310(99)00359-3

Gasparrini A, Armstrong B (2011) The impact of heat waves on mortality. Epidemiology 22:68–73. https://doi.org/10.1097/EDE.0b013e3181fdcd99

Gasparrini A, Guo Y, Sera F, Vicedo-Cabrera AM, Huber V, Tong S, Coelho MSZS, Saldiva PHN, Lavigne E, Correa PM (2017) Projections of temperature-related excess mortality under climate change scenarios. Lancet Planet Health 1:e360–e367. https://doi.org/10.1016/S2542-5196(17)30156-0

Gohar LK, Lowe JA, Bernie D (2017) The impact of bias correction and model selection on passing temperature thresholds. J Geophys Res (Atmos) 122:12045–12061. https://doi.org/10.1002/2017JD026797

Guo Y, Gasparrini A, Armstrong BG, Tawatsupa B, Tobias A, Lavigne E, Coelho MSZS, Pan X, Kim H, Hashizume M (2017) Heat wave and mortality: a multicountry, multicommunity study. Environ Health Perspect 125:087006. https://doi.org/10.1289/EHP1026

Hajat S, Kosatky T (2010) Heat-related mortality: a review and exploration of heterogeneity. J Epidemiol Community Health 64:753–760. https://doi.org/10.1136/jech.2009.087999

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York

Hastie T, Tibshirani R, Wainwright M (2015) Statistical learning with sparsity: the lasso and generalizations. CRC Press, Boca Raton

He H, Ma Y (2013) Imbalanced learning: foundations, algorithms, and applications. John Wiley & Sons, Hoboken

Hendriks C, Forsell N, Kiesewetter G, Schaap M, Schoepp W (2016) Ozone concentrations and damage for realistic future European climate and air quality scenarios. Atmos Environ 144:208–219. https://doi.org/10.1016/j.atmosenv.2016.08.026

Hersbach H, Dee D (2016) ERA5 reanalysis is in production. ECMWF Newsletter 147:7. Available at: https://www.ecmwf.int/en/newsletter/147/news/era5-reanalysis-production. Accessed 27 Apr 2020

Hertig E (2020) Health-relevant ground-level ozone and temperature events under future climate change using the example of Bavaria, Southern Germany. Air Qual Atmos Health 13:435–446. https://doi.org/10.1007/s11869-020-00811-z

Hertig E, Schneider A, Peters A, von Scheidt W, Kuch B, Meisinger C (2019) Association of ground-level ozone, meteorological factors and weather types with daily myocardial infarction frequencies in Augsburg, Southern Germany. Atmos Environ 217:116975. https://doi.org/10.1016/j.atmosenv.2019.116975

Husar RB, Renard WP (1998) Ozone as a function of local wind speed and direction: evidence of local and regional transport. Center for Air Pollution Impact and Trend Analysis, St Louis

Intergovernmental Panel on Climate Change (IPCC) (2013) Climate change 2013: the physical science basis. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge and New York, pp 1535

Jacob D, Petersen J, Eggert B, Alias A, Christensen OB, Bouwer LM, Braun A, Colette A, Déqué M, Georgievski G (2014) EURO-CORDEX: new high-resolution climate change projections for European impact research. Reg Environ Change 14:563–578. https://doi.org/10.1007/s10113-013-0499-2

Jacob DJ, Winner DA (2009) Effect of climate change on air quality. Atmos Environ 43:51–63. https://doi.org/10.1016/j.atmosenv.2008.09.051

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning - with applications in R. Springer, New York

Katragkou E, Zanis P, Kioutsioukis I, Tegoulias I, Melas D, Krüger BC, Coppola E (2011) Future climate change impacts on summer surface ozone from regional climate-air quality simulations over Europe. J Geophys Res (Atmos) 116:D22307. https://doi.org/10.1029/2011JD015899

Katsouyanni K, Analitis A (2009) Investigating the synergistic effects between meteorological variables and air pollutants: results from the European PHEWE, EUROHEAT and CIRCE projects. Epidemiology 20:S264. https://doi.org/10.1097/01.ede.0000362883.27030.8f

Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. Int J Clim 22:1441–1453. https://doi.org/10.1002/joc.773

Krueger O, Hegerl GC, Tett SFB (2015) Evaluation of mechanisms of hot and cold days in climate models over Central Europe. Environ Res Lett 10:014002. https://doi.org/10.1088/1748-9326/10/1/014002

Matthies F, Bickler G, Marín NC (2008) Heat-health action plans: guidance. World Health Organization, Copenhagen

McGregor GR, Bessmoulin P, Ebi K, Menne B (2015) Heatwaves and health: guidance on warning-system development. WMOP, Geneva

Meehl GA, Tebaldi C (2004) More intense, more frequent, and longer lasting heat waves in the 21st century. Science 305:994–997. https://doi.org/10.1126/science.1098704

Meehl GA, Tebaldi C, Tilmes S, Lamarque J-F, Bates S, Pendergrass A, Lombardozzi D (2018) Future heat waves and surface ozone. Environ Res Lett 13:064004. https://doi.org/10.1088/1748-9326/aabcdc

Moghani M, Archer CL (2020) The impact of emissions and climate change on future ozone concentrations in the USA. Air Qual Atmos Health. https://doi.org/10.1007/s11869-020-00900-z

Nuvolone D, Petri D, Voller F (2018) The effects of ozone on human health. Environ Sci Pollut Res 25:8074–8088. https://doi.org/10.1007/s11356-017-9239-3

Oswald EM, Dupigny-Giroux L-A, Leibensperger EM, Poirot R, Merrell J (2015) Climate controls on air quality in the Northeastern U.S.: an examination of summertime ozone statistics during 1993–2012. Atmos Environ 112:278–288. https://doi.org/10.1016/j.atmosenv.2015.04.019

Otero N, Sillmann J, Schnell JL, Rust HW, Butler T (2016) Synoptic and meteorological drivers of extreme ozone concentrations over Europe. Environ Res Lett 11:024005. https://doi.org/10.1088/1748-9326/11/2/024005

Pattenden S, Armstrong B, Milojevic A, Heal MR, Chalabi Z, Doherty R, Barratt B, Kovats RS, Wilkinson P (2010) Ozone, heat and mortality: acute effects in 15 British conurbations. Occup Environ Med 67:699–707. https://doi.org/10.1136/oem.2009.051714

Peng C-YJ, Lee KL, Ingersoll GM (2010) An introduction to logistic regression analysis and reporting. J Educ Res 96:3–14. https://doi.org/10.1080/00220670209598786

Robine JM, Michel JP, Herrmann FR (2012) Excess male mortality and age-specific mortality trajectories under different mortality conditions: a lesson from the heat wave of summer 2003. Mech Ageing Dev 133:378–386. https://doi.org/10.1016/j.mad.2012.04.004

Schnell JL, Prather MJ (2017) Co-occurrence of extremes in surface ozone, particulate matter, and temperature over eastern North America. Proc Natl Acad Sci 114:2854–2859. https://doi.org/10.1073/pnas.1614453114

Schoetter R, Cattiaux J, Douville H (2015) Changes of western European heat wave characteristics projected by the CMIP5 ensemble. Clim Dyn 45:1601–1616. https://doi.org/10.1007/s00382-014-2434-8

Sillman S, Samson PJ (1995) Impact of temperature on oxidant photochemistry in urban, polluted rural and remote environments. J Geophys Res (Atmos) 100:11497–11508. https://doi.org/10.1029/94JD02146

Song X, Wang S, Hu Y, Yue M, Zhang T, Liu Y, Tian J, Shang K (2017) Impact of ambient temperature on morbidity and mortality: an overview of reviews. Sci Total Environ 586:241–254. https://doi.org/10.1016/j.scitotenv.2017.01.212

Srebot V, Gianicolo EAL, Rainaldi G, Trivella MG, Sicari R (2009) Ozone and cardiovascular injury. Cardiovasc Ultrasound 7:30. https://doi.org/10.1186/1476-7120-7-30

Takahashi K, Honda Y, Emori S (2007) Assessing mortality risk from heat stress due to global warming. J Risk Res 10:339–354. https://doi.org/10.1080/13669870701217375

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Teutschbein C, Seibert J (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. J Hydrol 456-457:12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052

The Expert Team on Climate Change Detection and Indices (ETCCDI) (2009) Climate change indices. Definitions of the 27 core indices. Retrieved from: http://etccdi.pacificclimate.org/list_27_indices.shtml. Accessed 24 May 2020

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J Roy Stat Soc Ser B (Method) 58:267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tibshirani R (2011) Regression shrinkage and selection via the lasso: a retrospective. J Roy Stat Soc Ser B (Stat Method) 73:273–282. https://doi.org/10.1111/j.1467-9868.2011.00771.x

Wilks D (2006) Statistical Methods in the Atmospheric Sciences, 2nd edn. Academic Press, London

World Health Organization (WHO) (2006) WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide. World Health Organization, Geneva

World Health Organization (WHO) (2008) Improving public health responses to extreme weather. World Health Organization, Copenhagen

World Health Organization (WHO) (2013) Review of evidence on health aspects of air pollution–REVIHAAP Project. World Health Organization, Copenhagen

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.