Multivariate model-based investigation of the temperature dependence of ozone concentration in Finnish boreal forest

Isokääntä, Sini

2022-11-15

Isokääntä , S , Mikkonen , S , Laurikainen , M , Buchholz , A , Schobesberger , S , Blande , J D , Nieminen , T , Ylivinkka , I , Bäck , J , Petäjä , T , Kulmala , M & Yli-Juuti , T 2022 , ' Multivariate model-based investigation of the temperature dependence of ozone concentration in Finnish boreal forest ' , Atmospheric Environment , vol. 289 , 119315 . https://doi.org/10.1016/j.atmo

http://hdl.handle.net/10138/349857
https://doi.org/10.1016/j.atmosenv.2022.119315

Atmospheric Environment 289 (2022) 119315

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Multivariate model-based investigation of the temperature dependence of ozone concentration in Finnish boreal forest

Sini Isokääntä [a,*], Santtu Mikkonen [a,b], Maria Laurikainen [a], Angela Buchholz [a], Siegfried Schobesberger [a], James D. Blande [b], Tuomo Nieminen [c,d], Ilona Ylivinkka [c,e], Jaana Bäck [d], Tuukka Petäjä [c], Markku Kulmala [c], Taina Yli-Juuti [a]

[a] Department of Applied Physics, University of Eastern Finland, Kuopio, 70210, Finland
[b] Department of Environmental and Biological Sciences, University of Eastern Finland, Kuopio, 70210, Finland
[c] Institute for Atmospheric and Earth System Research (INAR) / Physics, Faculty of Science, University of Helsinki, Helsinki, 00014, Finland
[d] Institute for Atmospheric and Earth System Research (INAR) / Forest Sciences, Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, 00014, Finland
[e] Station for Measuring Ecosystem - Atmosphere Relations II (SMEAR II), University of Helsinki, Korkeakoski, 35500, Finland

HIGHLIGHTS

- Surface data differentiated apparent and direct temperature dependence of $[O_3]$.
- Volatile organic compounds contribute vitally to $[O_3]$ in this low-NOx environment.
- Statistical model reproduces the $[O_3]$, without exactly stating all involved processes.

ARTICLE INFO

Keywords:
Tropospheric ozone
Statistical modelling
Temperature dependence
OVOCs

ABSTRACT

Tropospheric ozone (O_3) concentrations are observed to increase with temperature in urban and rural locations. We investigated the apparent temperature dependency of daytime ozone concentration in the Finnish boreal forest in summertime based on long-term measurements. We used statistical mixed effects models to separate the direct effects of temperature from other factors influencing this dependency, such as weather conditions, long-range transport of precursors, and concentration of various hydrocarbons. The apparent temperature dependency of 1.16 ppb $^{\circ}C^{-1}$ based on a simple linear regression was reduced to 0.87 ppb $^{\circ}C^{-1}$ within the canopy for summer daytime data after considering these factors. In addition, our results indicated that small oxygenated volatile organic compounds may play an important role in the temperature dependence of O_3 concentrations in this dataset from a low-NOx environment. Summertime observations and daytime data were selected for this analysis to focus on an environment that is significantly affected by biogenic emissions. Despite limitations due to selection of the data, these results highlight the importance of considering factors contributing to the apparent temperature dependence of the O_3 concentration. In addition, our results show that a mixed effects model achieves relatively good predictive accuracy for this dataset without explicitly calculating all processes involved in O_3 formation and removal.

1. Introduction

Ground level, or tropospheric, ozone (O_3) is a harmful air pollutant for both humans and vegetation (e.g. Klingberg et al., 2011; Sousa et al., 2013), and it also plays a central role in the oxidative chemistry in the troposphere. Therefore, changes or trends in tropospheric O_3 concentration, such as temperature dependence, are of interest for many topics related to the atmosphere. Tropospheric O_3 concentration is affected by multiple factors with both anthropogenic and biogenic origin (Monks et al., 2015). The photolysis of NO_2 is the direct source for tropospheric O_3. Higher concentrations of nitrogen oxides ($NOx \equiv NO + NO_2$) and increased solar radiation will thus increase the observed O_3 concentration. Furthermore, any process which converts NO to NO_2 - such as the oxidation of carbon monoxide (CO), methane (CH_4) and other

* Corresponding author.
E-mail address: sini.isokaanta@uef.fi (S. Isokääntä).

https://doi.org/10.1016/j.atmosenv.2022.119315
Received 15 October 2021; Received in revised form 29 July 2022; Accepted 1 August 2022
Available online 6 August 2022

hydrocarbons (often also referred to as non-methane volatile organic compounds) by hydroxyl radicals - will increase O_3 production (e.g. Seinfeld et al., 2016). O_3 can also be transported from the stratosphere to the troposphere. Removal of O_3 can occur, for example, via dry deposition or various chemical reactions (Monks, 2005). Even though the mechanisms of O_3 production are rather well understood, linking atmospheric observations to specific chemical processes remains challenging due to the complicated interplay of O_3 formation and destruction mechanisms. Ozone has been intensively studied for multiple decades and decreases in peak O_3 concentrations have been observed for many locations across the globe (Monks et al., 2015; Yan et al., 2018). These decreases have been linked to successful regional emission controls, especially regarding anthropogenic sources of NOx and volatile organic compounds (VOCs). A similar decrease of peak O_3 concentration has also been observed in Finland for both urban and rural locations, but the trends for average O_3 concentration have been inconsistent (Anttila, 2020). Several locations in Europe have had increasing trends for the average O_3 concentration prior to the year 2000, but no positive trend later than that (Cooper et al., 2014). However, in some polluted areas, like Chinese megacities, the recent O_3 concentrations have been increasing despite the decreasing NOx levels (e.g., Li et al., 2021b).

The connections between O_3 concentration and temperature have been intensively studied (e.g., Sillman et al., 1995; Bloomer et al., 2009; Steiner et al., 2010; Rasmussen et al., 2013; Romer et al., 2018; Laban et al., 2020), and O_3 concentrations are observed to increase with temperature in multiple locations. For example, based on the review by Pusede et al. (2015), slopes between ~1 and 8 (ppb O_3 $^\circ C^{-1}$) have been reported for daily maximum O_3 concentration in summertime in various locations across the United States. The strongest effects have been observed in urban locations with relatively high NOx concentrations, but successful emission controls lowering the NOx levels and anthropogenic VOCs (i.e. O_3 precursors) have resulted in smaller temperature dependencies (Monks et al., 2015; Pusede et al., 2015). Indeed, the O_3 production rate response to changes in NOx levels differs depending on whether the O_3 production is taking place in a low-NOx or high-NOx regime (NOx or VOC limited O_3 production) (Seinfeld et al., 2016) making it important to distinguish investigated areas based on background NOx and VOC concentrations. Due to the complex chemistry leading to O_3 production, modelling studies with chemical transport models have also observed varying dependences between the resulting O_3 concentrations and temperature when emissions of biogenic VOCs (BVOCs) and their influence on O_3 formation reactions are considered (e.g., Ito et al., 2009). In addition, unresolved issues regarding O_3 sinks still remain (e.g. Zhou et al., 2017). Fewer studies exist covering low enough NOx levels corresponding to remote and rural regions. Coates et al. (2016) investigated the temperature dependency of O_3 in chemical transport model simulations, covering peak NOx concentrations between 0.02 and 10 ppb. They assigned the temperature dependent increase in O_3 concentration to be either caused by the chemistry itself (i.e. changes in reaction rates) or by the increase in isoprene emissions due to temperature. Romer et al. (2018), on the other hand, investigated observational data from a rural site at Centerville, Alabama in the southeastern United States, where they observed an increase in O_3 concentration of 2.3 ± 1 ppb $^\circ C^{-1}$ (1.7 ± 0.2 ppb $^\circ C^{-1}$ for long-term) with afternoon average NOx and isoprene concentrations of 0.3 ppb and 5.5 ppb, respectively, for the 2013 measurement campaign.

To interpret the observed apparent temperature dependence of the O_3 concentrations, multiple factors involved in the production and loss of O_3 and their response to temperature changes need to be considered. This includes, e.g., the concentration of NO which determines if a region is a source or a sink for O_3 (Seinfeld et al., 2016) or the concentration and reactivity of VOCs which can be oxidized by OH (forming OVOCs) and thus work as O_3 precursors. However, VOCs can also react directly with O_3, thus acting as a sink (Loreto et al., 2001). Furthermore, BVOC emissions from plants show a strong dependence on temperature. For example, monoterpene emissions increase exponentially with temperature (e.g. Guenther et al., 1995; Aalto et al., 2015; Hellen et al., 2018), thus possibly introducing a temperature dependence to the O_3 formation. Weather conditions (T, RH, wind, and solar radiation) and soil properties (soil T and moisture, affecting plant dynamics) also affect the local O_3 concentration in various ways (e.g. Ooka et al., 2011; Austin et al., 2015; Urban et al., 2017; Clifton et al., 2020b). Romer et al. (2018) deduced that for their study conditions, the increase in NOx emissions from microbial activity in the soil accounted for approx. 40% of the observed apparent temperature dependence of the O_3 concentration. The remaining change could be attributed to an increase in HOx (hydrogen oxide radicals) production with increasing global radiation (which is correlated with T). Laban et al. (2020), attempted to separate the effect of temperature on the daily maximum O_3 concentration from other variables, including changes in RH and NOx, for sites located in continental Africa with high anthropogenic and natural emissions. After accounting for these other changes, the increase of O_3 concentration due to temperature in their study is clearly smaller (~0.4–1.5 ppb $^\circ C^{-1}$) than reported earlier for urban locations (Pusede et al., 2015). However, both these studies (Romer et al., 2018; Laban et al., 2020) only covered short measurement campaigns of up to two years of continuous observations and could not investigate longer term trends.

In this study, we aim to determine the factors contributing to the apparent temperature dependence of O_3 concentration and to estimate the underlying change in O_3 concentration due to temperature at the Station for Measuring Ecosystem – Atmosphere Relations (SMEAR II) measurement site (Hari et al., 2005) in Hyytiälä, Finland. For this data-based investigation, we apply multivariate mixed effects models on the comprehensive long-term observations from the measurement site. The advantage of using a mixed effects model is that it enables the prediction of the O_3 concentration based on multiple processes without explicitly calculating all these processes. The study focuses on the summer months over a 15-year period. This station represents rural boreal forests, which account for almost a third of the forests around the globe (Fao, 2020). Due to the location of the site, SMEAR II is characterized by little urban pollution, the closest city being Tampere (238 140 inhabitants, Statistics Finland, 2019), located approximately 50 km from the measurement station. The station is surrounded by rather homogenous pine (*Pinus sylvestris*) forest, with monoterpenes being the largest group of BVOCs emitted (Rantala et al., 2015; Hellen et al., 2018).

2. Methods

2.1. Data selection and pre-processing

2.1.1. Measurements at SMEAR II, Hyytiälä, Finland

The SMEAR II (Station for Measuring Ecosystem–Atmosphere Relations: Hari et al., 2005) station provides long-term time series for multiple variables describing aerosols, gases, meteorology, soil, and plants at the site freely available to the public in an online database (Junninen et al., 2009). In this study we use statistical methods to investigate this data and in the final mixed effects regression model, we have used data from 2012 to 2018/2019. This period was chosen due to the continuous methane (CH_4) measurements that started in late 2012 and the end of the studied time period is dictated by the availability of VOC measurements. However, for additional analysis supporting our results, data since 2005 have also been investigated.

The atmospheric composition at SMEAR II is different between the different seasons. Winters usually see stronger anthropogenic influence, e.g. due to residential heating, whereas the emissions of BVOCs are more important during spring and summertime (e.g. Patokoski et al., 2014; Patokoski et al., 2015). To investigate O_3 concentrations in an environment that is strongly affected by biogenic emissions, we have used only summertime (June, July, and August) data. In addition, due to the very different chemistry of O_3 during day- and nighttime, we used only daytime data by selecting time points where the sun is at least one

degree above the horizon in Hyytiälä (lat. 61.80, lon. 24.30). This selection is consistent with previous studies investigating the maximum O_3 concentrations which do take place in daytime. The varying boundary layer height during the selected period of each day was taken into account as a variable in the further data analysis. Depending on the measured variable, the original time resolution varied. Thus, we averaged all data into 1-hourly values by taking the arithmetic mean for every hour. Due to the operation of the VOC measurements system, VOC concentrations are provided for every 3rd hour (having on average 9 observations within that hour), thus limiting our final investigation to one data point every 3 h (i.e., 1-h means for every 3rd hour). Data rows which had wind direction between 120° and 140° were removed from all the analyses to exclude the influence of the emissions from a nearby sawmill (Liao et al., 2011). Tables S1 and S2 show the data coverage relative to the hours available during daytime at summer after excluding the possible sawmill emissions.

We investigated the following variables (in addition to the O_3 concentration itself) for their potential effect on [O_3] at SMEAR II: wind speed (WS) and direction (WD), concentrations of nitrogen oxides (NOx, including NO and NO_2), equivalent black carbon concentration (eBC), carbon monoxide (CO) and CH_4 concentrations, atmospheric pressure (p), temperature (T), relative humidity (RH), global radiation (GlobR), soil temperature (soil T), soil moisture content (soil H_2O) and VOC concentrations. The VOCs include aromatics (benzene and toluene + p-cymene), biogenic VOCs (BVOCs; isoprene, and monoterpenes) and oxygenated VOCs (OVOCs; methanol, acetaldehyde, ethanol + formic acid, acetone, acetic acid, methacrolein + methyl vinyl ketone, and methyl ethyl ketone). The variables listed here were inspected as they are either known and/or proposed to affect the O_3 concentration directly or via pathways involving multiple processes. For example, CO, NOx and CH_4 are involved in the gas phase chemistry of O_3 formation and destruction (e.g., Seinfeld et al., 2016), RH has a strong connection to O_3 via controlling dry deposition (Altimir et al., 2006; Zhou et al., 2017), and VOCs can act as sources or sinks for O_3 (e.g., Vermeuel et al., 2021). In addition, the O_3 fluxes measured at SMEAR II were inspected for their potential dependence on temperature. Summary statistics and details of the measurement instruments are listed in Tables S3–S5.

Global radiation was used to calculate the clear sky parameter by dividing the measured global radiation by the calculated theoretical maximum radiation. This parameter describes the cloudiness at SMEAR II, with less clouds with increasing numerical value (Baranizadeh et al., 2014; Dada et al., 2017). In addition to the data measured within the forest canopy (measured at height 16.8 m), we also investigated (if available) the same observations from the measurement mast at 67.2 m thus being clearly above the ~18 m tall canopy.

2.1.2. Back-trajectory data and airmass source analysis

Hourly 96-h back trajectories were obtained from the HYSPLIT4 model (Hybrid Single-Particle Lagrangian Integrated Trajectory, Stein et al., 2015) for the same time period as the measurements from SMEAR II. Due to the relatively short lifetime of O_3 (Monks et al., 2015), we limited the trajectories to 24 h before reaching the station before clustering them into more specific airmass source areas. Longer trajectories (48–96 h) were tested, but as the airmass sources stayed approximately the same, shorter trajectories were used as they provided the most precise clustering results. The Euclidean distance matrix (Rencher et al., 2012) calculated from coordinates of the 24-h long airmass trajectories corresponding to the final model data were clustered with partitioning around medoids (PAM). The algorithm for PAM is fully described elsewhere (Kaufman et al., 1990). Six clusters were reasonably interpretable and clearly distinct. Thus, they were used as random effects in the regression models. A short description of the trajectory calculations, PAM and selection of the clusters along with their characteristic properties can be found in the SI.

The arrival height at Hyytiälä was set to 100 m above ground. In this study, the trajectories are included to consider possible variation in the local variable concentrations in Hyytiälä due to long-range transport of various pollutants or their precursors. In addition, as various output parameters along the trajectories can be obtained, we used the mixing layer height (MLH, km) 1 h before the airmass arrives at Hyytiälä to represent the mixing layer height at Hyytiälä due to the absence of actual measurements of this variable prior to 2013 at the site.

2.2. Multivariate mixed effects models

2.2.1. General description of mixed effects models

A multivariate mixed effects model was used to estimate the O_3 concentrations in Hyytiälä. A mixed effects model was used as these models estimate the variance-covariance structure of the data in addition to the mean of the response variable, and do not require standard homogeneity and independency assumptions (e.g.,Mcculloch et al., 2008; Demidenko, 2013) which are not often met with atmospheric data. Mixed effects models are also better justified than fixed effects models for grouped data sets with possible hierarchical structures (in this study, e.g. by airmass sources or hour of the day) (Demidenko, 2013; Mehtätalo et al., 2020). In addition, with mixed effects models, we can maintain the hourly resolution of the data and include the diurnal variation of the variables as a random effects without the need to average the data to efface the diurnal variability, similarly as in Mikkonen et al. (2020).

The linear mixed effects model can be presented in general form as

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\boldsymbol{y}$ is the vector of the response variable (here O_3 concentration), $\boldsymbol{\beta}$ (n x 1, n is the number of fixed effects) and $\boldsymbol{b}$ (q x 1, q is the number of random effects) are the vectors of fixed and random effects, respectively, and $\mathbf{X}$ (p x n, p is the number of observations) and $\mathbf{Z}$ (n x q) are the related design/coefficient matrices (Mcculloch et al., 2008) representing the predictor variables in the model (e.g. temperature, concentrations of various compounds, air mass source area, etc.). Vector $\boldsymbol{\varepsilon}$ includes the random errors. The formulation of the final regression model with the measured variables is presented in section 3.2 with Eq. (2). Depending on the random effect structure (crossed or nested effects), the relationship between $\mathbf{X}$ and $\mathbf{Z}$ varies (Mcculloch et al., 2008). The general idea of the random terms is to consider the group specific properties in the data, e.g., here airmass source specific effects on O_3 concentrations. More particularly, heterogeneity and correlation structure between and within the random effects of the data can be considered with justified choices of the design matrix $\mathbf{Z}$, and covariance structures for random parameters $\mathrm{Cov}(\boldsymbol{b}) = \mathbf{G}$ (q x q) and for residuals $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathbf{R}$ (n x n) can be defined and fitted. Modelling variances and covariances of the observations helps to provide valid statistical inference for the fixed effects $\boldsymbol{\beta}$ of the mixed model. As a difference to general linear models, the error terms $\boldsymbol{\varepsilon}$ can be correlated and the correlation structure can be defined, which makes the modelling more robust and accurate. Thus, the distribution of predicted observations $\hat{y}$ can be described by a distribution with the expectation of $\mathbf{X}\boldsymbol{\beta}$ and the covariance matrix $\mathbf{V}$ (n x n), which is given by $\mathbf{V} = \mathbf{ZGZ'}+\mathbf{R}$. However, as with multivariate regression models in general, the correlation structure within the independent variables (fixed effects) must be considered separately in order to avoid multicollinearity (e.g., Freund et al., 2006). More details of the content and construction of the presented matrices in the mixed effects model can be found from the literature (e.g.,Mcculloch et al., 2008; Demidenko, 2013; Zuur et al., 2009; Burzykowski et al., 2013).

A mixed effects model enables the prediction of the O_3 concentration based on multiple affecting factors (expressed by the variable groups) without explicitly calculating all the processes involved. While the model does not include explicit description of the processes and therefore specific processes cannot be investigated in detail, the role of various processes or variables in the changes in O_3 concentration can be studied. It is a computationally efficient way to find and quantify the

dependencies in the data with high accuracy. For example, Yli-Juuti et al. (2021) used a mixed effects model to estimate the effect of temperature on the organic aerosol mass loading. The linear mixed effects model was calculated with R statistical software (R 4.1.1, R Core Team, 2021) with the *lmer*-function provided in the package lme4 (Bates et al., 2015). The restricted maximum likelihood method (e.g.,Mcculloch et al., 2008) was used for estimating the model parameters.

2.2.2. Selection of relevant variables for the mixed effects model

In this section we justify our decision to leave out a few variables, which are usually considered relevant for O_3 formation, from the final mixed effects regression model. All the available variables listed in Table S1 were investigated for their impact on the concentration of O_3. First, the model Bayesian information criterion (BIC) (Schwarz, 1978) was inspected to see if the added variable increased the overall fit of the model. Second, the significance of a certain variable (a significance level of $\alpha = 0.05$ was used) to the model was determined by likelihood ratio tests (Wilks, 1938) calculated between different model versions. In addition to comparing two nested model structures, we compared non-nested models which had the same number of degrees of freedom (e.g., $model_1$ with the predictors T, [CO], [BC] and [NO] and $model_2$ with T, [CO], [BC] and [NOx]). Due to the non-nested structure, the likelihood ratio test cannot be directly used to determine if one variable is a better predictor than another. Instead, we inspected the Pearson correlation coefficients (Pearson's r) between modelled and measured data and inspected the other regression coefficients (in addition to that for T), as large changes in the coefficients (when changing a predictor in a model into another) could indicate a variable has relevance in the model. The Pearson's r values were also inspected between the predictors in order to avoid multicollinearity.

The first exploratory investigation revealed that some variables had no significant effect (i.e., were not statistically significant based on the process explained above) or were not suitable due to their own dependence (Pearson's r above 0.5) on T or other predictor variables. Below, we further elaborate why these variables were omitted from additional analysis.

Even though RH is an important variable for O_3 formation through affecting O_3 loss via dry deposition, and it is a statistically significant predictor in the model, we have excluded it from the final model due to its high correlation with T (Pearson's r-0.56 for summer during daytime between 2005 and 19) to avoid multicollinearity. However, the effect of RH is also partly covered by the clear sky parameter as high RH corresponds to situations with more cloud cover at SMEAR II: the Pearson's r value between hourly RH and the clear sky parameter is −0.64 for summertime data for daylit hours between 2006 and 2019. This connection is likely partly due to the drop in temperature due to more cloud cover (e.g., Seinfeld et al., 2016). In addition, during the progression of the data analysis, we observed that the NOx and NO_2 concentrations were not significant predictors for O_3 concentration in this data set and are thus omitted from the regression models presented below. However, NO concentration was found to be significant instead. This is likely caused by NO being more directly linked to the formation of O_3 and thus also to the observed O_3 concentration (e.g., Seinfeld et al., 2016). NO_2, however, can both enhance and decrease the O_3 production and thus the model is not able to detect a significant net effect for $[O_3]$ due to $[NO_2]$. In addition, in SMEAR II $[NO] \ll [NO_2]$, thus small changes in NO concentration are not as pronounced in the total [NOx] which reduces the quality of [NOx] as a predictor. However, it has to be noted that the mixed effects model simply looks for predictors that are significant in the context of its framework. For a detailed chemical box model, which investigates the actual chemical processes, both the NO and NO_2 concentrations would be relevant. The measured OVOC concentration is highly dependent on temperature, leading to an exponential increase of OVOC concentration as a function of T (Fig. S1). Thus, OVOC concentration also has a high correlation with temperature (Pearson's r = 0.65 for daytime between 2010 and 18), thus preventing their use in the regression model together with temperature. We investigated their relevance for O_3 concentration by other means (see section 3.4). Other variables excluded from the regression are concentration of aromatics, pressure, and soil T. These were not statistically significant predictors in the models.

The random effects investigated are hour of the day, airmass source, and observation year. Variations in O_3 concentration by these effects is visualized in Fig. S2. Hour of the day is needed in the regression model to include diurnal variation not captured by the other variables in the model. The use of airmass sources is justified due to differences in O_3 concentration between these sources. The observation year is used to consider any possible year-by-year variation in the measured data caused by factors not directly considered with the fixed variables in the model, such as instrumental reasons (e.g., the device measuring [eBC] changed in 2017, causing a small shift in the measured eBC concentrations).

To investigate how the effect of temperature changes as more factors affecting the O_3 concentration in the troposphere are taken into account, we separated the model predictors into different groups: basic chemistry, anthropogenic influence, weather, BVOCs, soil/plant processes and random effects. Variables included in each group are shown in Table 1. The role of OVOCs was investigated separately (see section 3.4). The BVOCs represents the sum of MTs (monoterpenes) and isoprene concentrations, as the effect of those VOCs on O_3 was in the same direction with similar magnitude. Combining these VOCs did not affect the observed T dependency outside of the error limits. Regarding the weather variables, we observed that the wind speed (WS) is highly affected by the wind direction (WD), thus the WD was categorised into east (E, 45°–120°), south (S, 140°–225°), north (N, 0°–45° and 315°–365°) and west (W, 225°–315°) and we allowed the fixed effect for WS to vary depending on the categorised wind direction. Using more (e.g., 8) categories for the WD neither improved the overall fit nor changed the regression coefficients (including that for T) in the final multivariate mixed effects model, thus using 4 categories is sufficient. The mathematical formulation of the complete model with all the predictors is shown in Eq. (2) in section 3.2. All statistical analysis in this study was conducted with R statistical software (R 4.1.1, R Core Team, 2021).

3. Results and discussion

3.1. Results from simple regression

For comparison to existing literature on O_3 concentration dependence with temperature, simple regression fits were first calculated for 1-h resolution data. Long-term observations show an increase in O_3 concentration with temperature (Fig. 1a). The slope of $[O_3] \sim T$ from the ordinary least squares (OLS; Chambers, 1992) regression estimates this increase to be 1.20 ppb $°C^{-1}$ (Pearson's r = 0.62 between modelled and predicted $[O_3]$), whereas a Deming regression (Deming, 1943) suggests

Table 1
Model predictor groups and variables in each group.

Group	Included variables
Basic chemistry	NO, CO, CH_4
Anthropogenic influence	eBC
Weather	wind speed and direction, cloudiness (clear sky-parameter) and MLH
BVOCs	MTs and isoprene (sum of the concentrations)
Soil/plant processes	Soil H_2O
OVOCs*	methanol, acetaldehyde, ethanol + formic acid, acetone, acetic acid, methacrolein + methyl vinyl ketone, and methyl ethyl ketone (sum of the concentrations)
Random effects	Hour of the day, airmass origin (cluster), and observation year

* This group is excluded from the final model (section 3.2) but assessed in section 3.4.

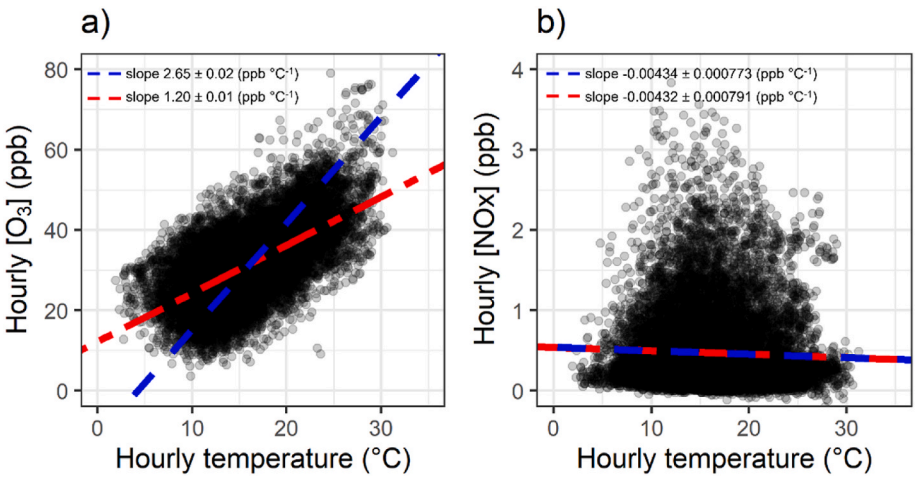

Fig. 1. Daytime (sun 1° above the horizon) hourly a) O_3 and b) NOx concentrations as a function of temperature for June–August 2005–2019 at SMEAR II, Hyytiälä, Finland within the forest canopy. The blue line shows a Deming regression fit and red line OLS (ordinary least squares regression) fit for the [O_3] and [NOx] as a function of ambient T. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

an increase of 2.65 ppb °C^{-1} (Pearson's r = 0.62 between modelled and measured [O_3]). The difference between the estimates is surprisingly large, but generally methods considering the errors for both x and y variables (i.e. as in Deming regression) should be preferred for this type of bivariate analysis (Mikkonen et al., 2019). Measured O_3 fluxes do not show significant trends as a function of temperature (SI Section S1 and Fig. S3). The NOx levels in Hyytiälä (daytime median 0.30 ppb between 2005 and 2019) are similar to those reported in Romer et al. (2018) for the rural US (slope of 1.7–2.3 ppb °C^{-1} between O_3 and temperature) and for both sites no apparent change in the NOx concentrations (slope from OLS regression −0.0043 ppb °C^{-1} with Pearson's r = 0.04 between modelled and measured [NOx] at Hyytiälä) was observed with increasing temperature (Fig. 1b). However, at Hyytiälä the highest NOx concentrations for this dataset ([NOx] > 2 ppb) are mostly observed at T between 10 °C and 25 °C. The main difference between the sites (SMEAR II and the site in rural US in Romer et al., 2018) is the type of the surrounding forest, leading to more than one order of magnitude lower isoprene concentration at SMEAR II (daytime median of 0.12 ppb between 2010 and 2019).

Relatively high O_3 concentrations (summertime daytime median 31 ppb between 2005 and 19) were observed despite the rather low NOx concentration (summertime daytime median 0.30 ppb between 2005 and 19). This could be due to low/missing sinks for O_3. For example, Li et al. (2021a) showed that the typical day-night cyclicity is much less prominent at SMEAR II compared to the Landes forest (pine forest in France), where much stronger O_3 loss by reactions with MTs can be seen. However, at the Landes forest site, the MT concentration is ~10 times higher than at SMEAR II (Li et al., 2021a). Vermeuel et al. (2021) also proposed, that in-canopy chemistry of BVOCs plays a significant role in the removal of O_3, suggesting that the O_3 sinks are weaker in SMEAR II where the BVOC concentrations are much smaller. Generally, BVOCs are suggested to play a significant role in removal of O_3 (Wolfe et al., 2011). However, Zhou et al. (2017) estimated that the removal of O_3 by chemical processes is relatively small at SMEAR II, compared to, e.g., deposition on wet surfaces.

3.2. Estimating the temperature dependence of O_3 by the mixed effects model

We estimated the temperature dependence of the O_3 concentration from the fixed coefficient for temperature in the mixed effects model, which can be expressed as

$$[O_{3i}] = \beta_0 + \{\boldsymbol{b_h} + \boldsymbol{b_y} + \boldsymbol{b_a}\} + \beta_1 T_i + \{\beta_2[\mathrm{NO}_i] + \beta_3[CH_{4,i}] + \beta_4[CO_i]\} + \{\beta_5[BVOC_i]\} + \{\beta_6[eBC_i]\} + \{\boldsymbol{\beta_{WS:WD}} WS + \beta_7 ClearSky_i + \beta_8 MLH_i\} + \{\beta_9 Soil_{H_2O,\,i}\} \quad (2)$$

where β_0 is a model fixed intercept, $\boldsymbol{b_h}$, $\boldsymbol{b_y}$ and $\boldsymbol{b_a}$ are the vectors of random intercepts for hour of the day, year, and airmass source, respectively. β_1- β_9 are the fixed coefficients for T, [NO], [CH_4], [CO], [BVOC], [eBC], clear sky parameter, MLH, and soil H_2O, respectively. $\boldsymbol{\beta_{WS:WD}}$ is the (4x1) vector of coefficients for wind speed dependent on wind directions. Each variable group presented in Table 1 is separated with curly brackets in Eq. (2).

In addition to the final version of the model, shown in Eq. (2) and discussed more in Section 3.3, we analyzed different versions of the model where specific variables or groups of variables were added in the model one at a time. To achieve consistent results, variables shown in Table 1 (except OVOCs) were first merged into one dataset (containing both measurement heights) with unified time resolution, thus leading into a final dataset with 3-h resolution due to the limitations in the VOC measurements. Only periods, when all parameters were available, were included in the analysis, and all time points where one or more of the variables were not available, were removed prior to the regression.

Table 2 shows the results from the regression model as more groups (i.e., all variables within one group) are cumulatively added to the model. The Bayesian information criterion (BIC, Schwarz, 1978) and the correlation (Pearson's r) between the observed and predicted values describe the goodness-of-fit. The simplest model (OLS regression) is a simple univariate regression model which is only based on the air T (first row in Table 2). These values differ slightly from the slopes shown in Fig. 1 as the data set size has changed due to the merging of the selected variables. The OLS regression is included as a reference point to compare with previous studies using the same approach. We then evaluate how the successive addition of variable groups change the apparent T dependence of the O_3 concentration. The respective slopes for T

Table 2
Regression model evolution per added variable group. The variable groups were added cumulatively from top (random effects) to bottom (soil/plant processes). The model evolution is shown with BIC (Bayesian information criterion) and Pearson's r between observed and model predicted data. Data from both heights include simultaneous observations. The OLS fit is included to have a reference point to earlier studies.

Added variable group	Slope for T (ppb °C^{-1}) i.e. β_1 in Eq. (2)		BIC		Pearson's r (obs. vs pred. [O_3])	
Measurement height	within canopy	above canopy	within canopy	above canopy	within canopy	above canopy
OLS ([O_3] ~ T)	1.16 ± 0.04	1.12 ± 0.04	7777	7720	0.64	0.63
Base case:						
[O_3] ~ T + random effects	0.91 ± 0.04	0.90 ± 0.04	7546	7471	0.74	0.74
Variable group added to the base case (cumulatively):						
Basic chemistry	0.93 ± 0.05	0.88 ± 0.04	7493	7402	0.76	0.76
BVOCs	1.08 ± 0.05	0.98 ± 0.04	7363	7330	0.79	0.78
Anthropogenic influence	0.98 ± 0.04	0.85 ± 0.04	7290	7231	0.80	0.80
Weather	0.74 ± 0.05	0.63 ± 0.05	7153	7111	0.83	0.83
Soil/plant processes	0.87 ± 0.05	0.76 ± 0.05	7108	7063	0.84	0.84

(regression coefficients β_1) and goodness-of-fit values are shown in the rows following the OLS (including only T as a predictor variable) case in Table 2. For each case, temperature was included in the model and the variable groups were successively added in the order as they are listed. When we add a variable group in the mixed effects model, a part of the variability of the O_3 concentration can then be explained by the processes represented by these variables. This will separate the direct effect of temperature on the O_3 concentration from these other dependences without explicitly modelling them (e.g., all possible VOC reactions with O_3). A change in the regression coefficient β_1 (slope for T) when a new variable group is added, indicates that the previously calculated apparent T dependence included these other effects. Each added group improves the fit (decreasing BIC and increasing Pearson's r). Addition of the variable groups also increased the overall fit significantly based on likelihood ratio tests (Wilks, 1938). In addition, the changes in the apparent temperature dependence of the O_3 concentration when each group is added separately without the other ones is presented in Table 3.

Adding the random effects decreases the apparent T dependence by 0.25 ppb °C^{-1} within the canopy and by 0.22 ppb °C^{-1} above the canopy (Table 2). The addition of the basic chemistry components does not change the apparent T dependence much, but it improves the prediction of the O_3 concentration as indicated by the decreasing BIC and increasing Pearson's r. This could suggest none of the components in that group (basic chemistry) show significant changes with temperature and/or their concentrations are so high that small variations in them do not affect the predicted O_3 concentration. It is also possible that the effects of these components counteract each other, thus the net effect regarding the T dependence of [O_3] is close to zero. The largest decrease in the slope for the cumulative approach (Table 2) is caused by the variable group "weather". To further investigate if this large decrease in the temperature dependence is due to a single term within the weather parameters, we constructed additional models in which these parameters were investigated separately. Based on this analysis, the clear-sky parameter causes the largest decrease in the slope (for example, from 0.98 to 0.79 within canopy with BIC decrease from 7290 to 7251) but the wind terms and MLH improve the model more (BIC decreases from 7290 within canopy to 7207 or 7208, respectively).

Table 3
Regression model evolution per added variable group. The variable groups were added individually to the base case model including T and the random effects. The model evolution is shown with BIC (Bayesian information criterion) and Pearson's r between observed and model predicted data. Data from both heights include simultaneous observations.

	Slope for T (ppb °C^{-1}) i.e. β_1 in Eq. (2)		BIC		Pearson's r (obs. vs pred.)	
Measurement height	within canopy	above canopy	within canopy	above canopy	within canopy	above canopy
Base case:						
[O_3] ~ T + random effects	0.91 ± 0.04	0.90 ± 0.04	7546	7471	0.74	0.74
Variable group added to base case (individually):						
Basic chemistry	0.93 ± 0.05	0.88 ± 0.04	7493	7402	0.76	0.76
BVOCs	1.10 ± 0.04	1.03 ± 0.04	7405	7393	0.78	0.76
Anthropogenic influence	0.84 ± 0.05	0.77 ± 0.05	7539	7441	0.74	0.75
Weather	0.79 ± 0.04	0.78 ± 0.04	7349	7347	0.80	0.78
Soil/plant processes	0.99 ± 0.05	0.98 ± 0.04	7527	7450	0.75	0.75

3.3. Validating the temperature dependence of O_3

The final model shown in Eq. (2) corresponds to the last row in Table 2, which includes the factors found statistically significant for the observed [O_3] from the variables we tested (Tables S1 and S2) as described in Section 2.2.2. The predictor variables in this model also improved the overall fit significantly when added one-by-one instead of adding them as groups (not shown). The fixed regression coefficients and their significance levels for the final model are presented in Table 4 and the intercepts (fixed and random) are visualized in Figs. S8 and S9.

Table 4
Regression coefficients from the final mixed effects model fit. The * indicates variables that were observed also at the above canopy level, in addition to O_3.

Fixed effect/ slope	Slope ± Std. Error		P-value		Unit
	At canopy	Above canopy	At canopy	Above canopy	
T*	0.87 ± 0.05	0.76 ± 0.05	<0.001	<0.001	ppb/°C
NO*	−15.85 ± 3.58	−18.34 ± 3.06	<0.001	<0.001	ppb/ppb
CH_4*	−0.09 ± 0.01	−0.10 ± 0.01	<0.001	<0.001	ppb/ppb
CO*	0.08 ± 0.01	0.10 ± 0.01	<0.001	<0.001	ppb/ppb
Clearsky	3.75 ± 0.78	4.74 ± 0.74	<0.001	<0.001	ppb/1
MLH	2.61 ± 0.47	2.21 ± 0.43	<0.001	<0.001	ppb/km
eBC	16.69 ± 1.72	19.53 ± 1.67	<0.001	<0.001	ppb/ μgm^{-3}
BVOCs*	−4.72 ± 0.43	−5.63 ± 0.61	<0.001	<0.001	ppb/ppb
Soil H_2O	26.90 ± 3.80	27.44 ± 3.75	<0.001	<0.001	ppb/ m^3m^{-3}
WS*:WD(E)	0.50 ± 0.34	0.18 ± 0.12	0.1370	0.1180	ppb/ ms^{-1}
WS*:WD(N)	0.53 ± 0.35	0.16 ± 0.16	0.1330	0.3130	ppb/ ms^{-1}
WS*:WD(S)	1.64 ± 0.28	0.64 ± 0.11	<0.001	<0.001	ppb/ ms^{-1}
WS*:WD(W)	1.53 ± 0.25	0.56 ± 0.10	<0.001	<0.001	ppb/ ms^{-1}

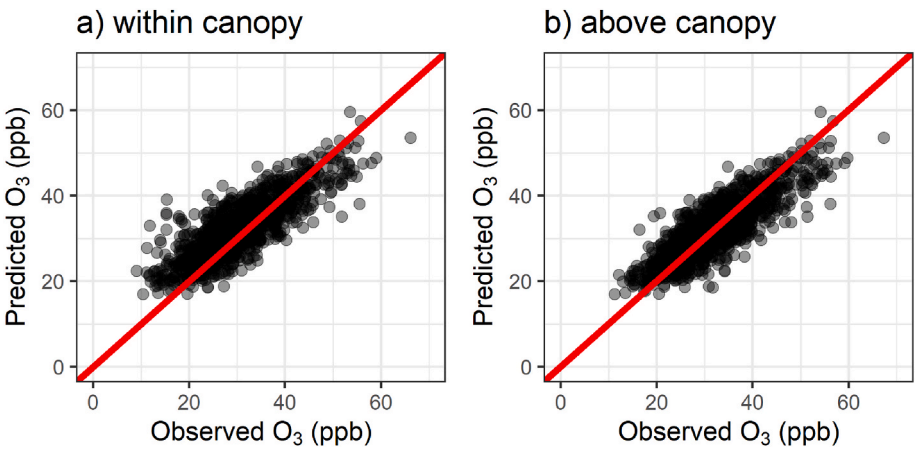

Fig. 2. Predicted vs. the observed O_3 concentration as a scatter plot a) within the canopy and b) above the canopy with the final model presented in Eq. (2). Pearson's r between the observed and predicted [O_3] for both heights is 0.84. The 1:1 line is shown in red as a reference for the "perfect fit". (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

The model yields a temperature dependence of 0.87 ppb O_3 $^\circ C^{-1}$ and 0.76 ppb O_3 $^\circ C^{-1}$ within canopy and above canopy, respectively, with rather good agreement between the observed and predicted O_3 concentration (Pearson's r 0.84 within and above canopy) as shown in Fig. 2 and reported in the last row of Table 2. The highest O_3 concentrations are slightly underestimated (Fig. 2) suggesting some processes are still missing. However, this is expected, as e.g., OVOCs known to affect O_3 (Mellouki et al., 2015) were not included. To further validate our regression model, we ran 1000 bootstrap replicates (e.g., Efron et al., 1986; see details from S4.2) to investigate if the model coefficients are sensitive to changes in the data or if significant bias exists. The distribution of the regression coefficients for T in those 1000 replicates are normally distributed (Fig. S10) and no clear biases were observed for the coefficients of the other variables either, indicating our regression model has reasonable capability to predict O_3 concentration at SMEAR II.

The obtained temperature dependence falls at the lower end of the values reported in previous studies for similar conditions. However, it needs to be considered that the mixed effects model separates the direct T dependence from the other factors. Studies that also distinguish multiple different factors show lower apparent T dependence values as well (e.g. Laban et al., 2020), when compared to bivariate models (Pusede et al., 2015 and references therein). The simplified box model analysis presented by Romer et al. (2018) suggested that 60% of the apparent T dependence can be attributed to an increase in the HOx production, which primarily depends on the global radiation but appears to be temperature dependent. With our approach, this effect would be mostly covered by the clear sky parameter (Pearson's r between clear sky parameter and GlobR in the final data is 0.80) and thus does not affect the apparent T dependence to the same degree.

The direct T dependence of the O_3 concentration is related to the temperature dependence of chemical reactions that are involved in producing O_3. Often, the reaction rates involving O_3 production are dependent on temperature and, in many cases, reactions involving oxidation of VOCs by OH, and NOx oxidation by HOx and ROx do actually have negative dependency on T (e.g., in MCMv3.2, Rickard et al., 2021). This would suggest a decrease of O_3 with increasing T. However, as the net O_3 production is not only determined by these oxidation reactions, the actual net change in O_3 concentration might still end up positive. For example, Coates et al. (2016) predicted an increase of O_3 with T due to changes in reaction rates to be approximately 0.25–0.52 ppb $^\circ C^{-1}$ with their chemical mechanism models under low-NOx conditions.

3.4. The role of OVOCs

Pusede et al. (2014) observed that the total organics reactivity with OH increases as a function of temperature. This increase is suggested to be driven by increasing concentrations of VOCs with temperature, most importantly of small oxygenated organic molecules. Even though we are not able to include both OVOCs and T in our model, we can investigate the effect of OVOCs with a similar approach. First, we created a dataset containing the OVOCs presented in Table 1 along with all the other variables used in the regression so far. As this slightly limited our dataset, we first refitted the regression model from section 3.2 with Eq. (2) (i.e., no OVOCs), which gave a slope of 0.87 ± 0.06 (ppb O_3 $^\circ C^{-1}$) for T within the canopy, with a Pearson's r of 0.848 between measured and predicted O_3. To investigate the possible effect of OVOCs, we then replaced the term $\beta_1 T_i$ (which captures the T dependence) with a term for the dependence of the O_3 concentration on the OVOCs concentration ($\beta_{10}[OVOC_i]$) and looked into the predictive capability of the model for the O_3 concentration. Using the OVOC concentration instead of temperature, resulted in a Pearson's r of 0.870, higher than for the model with temperature (0.848). The slope for [O_3] ~ [OVOCs] (β_{10}) was positive (1.23 ± 0.06 ppb O_3 ppb^{-1}) in contrast to the slope for BVOCs (−4.91 ± 0.45 ppb O_3 ppb^{-1}). Regression coefficients for the other variables and intercepts (fixed and random) are reported in the supplementary material (Figs. S11–S12 and Table S7). A negative slope for BVOCs could indicate that those compounds act as a net "sink" for O_3 due to their reaction with O_3, or that O_3 is produced if the dominant BVOC reaction was with OH. Most of the OVOCs, on the other hand, do not react with O_3 directly as they do not contain C=C double bonds. Those OVOCs (see Table 1) could be a source of O_3 via their oxidation by OH, but they can also be reaction products of BVOC oxidation and thus be an indication of the amount of reacted BVOCs. Therefore, the different signs for the slopes of BVOC and OVOC are reasonable, however it is not possible to conclude which is the dominant process. However, if the OVOCs were formed dominantly by the reaction of BVOCs with O_3, it might indicate an O_3 loss process, and we would expect a negative slope for OVOCs. Our methodology cannot resolve the actual processes involved here. However, our results show that further investigation of the relationship of OVOCs and O_3 are needed.

Our findings suggest that there is a rather strong, temperature dependent connection between [O_3] and [OVOCs]. Whether this is solely due to the increase of OVOC concentration or changes in their reactivity caused by temperature, as suggested by Pusede et al. (2014), is

hard to say. In addition, small-sized OVOCs increasing with T could also be connected to their reduced uptake on, or enhanced release from, wet surfaces upon the drying of these surfaces with increasing T (Fulgham et al., 2020).

To get a rough estimate for the temperature effect when taking the concentration of OVOCs into account, we selected acetaldehyde and ethanol + formic acid from the full set of available OVOCs to be included in the model with temperature. This was possible as the Pearson's r between T and individual OVOC concentrations varied, being smallest for those selected (Pearson's r values < 0.5), thus decreasing the risk of multicollinearity. With this setup, an $[O_3] \sim T$ slope of 0.67 ± 0.06 (ppb O_3 $^\circ C^{-1}$) was obtained with a Pearson's r of 0.876 between the observed and predicted $[O_3]$. This further supports the hypothesis that the OVOCs play an important role in the observed T dependence of O_3 concentration. This link between $[O_3]$ and [OVOCs] is included in the slope between $[O_3]$ and T obtained with simple linear regressions. This creates an illusion of a rather strong temperature dependence of O_3 concentration, even though part of this dependence may not be directly caused by temperature but rather e.g. due to the temperature dependence of other relevant factors such as possible nonlinear processes involving the chemistry of OVOCs.

The other variables in the model with OVOCs instead of T had only minor changes in their coefficients considering the changes in the data (Table S7), but soil H_2O became insignificant for the model. This could indicate some interaction between OVOCs and the soil H_2O (e.g., Fulgham et al., 2020), possibly linked to variations in plant and forest floor emissions (see e.g. Aaltonen et al., 2013; Mäki et al., 2019), as OVOCs can also be directly emitted by plants (e.g., Harley et al., 2007). However, plants can also take up OVOCs and thus the stomatal closure caused by drought could both decrease the emissions and reduce the uptake of OVOCs (Filella et al., 2009; Saunier et al., 2017). Lin et al. (2020) observed increasing O_3 levels with droughts. Thus, we might expect higher O_3 concentration for dryer soils also for Hyytiälä. The Pearson's r between O_3 and soil H_2O is slightly negative (-0.09 between 2005 and 19 and -0.33 for the final model data), indicating higher O_3 concentrations occur more often with dryer soils. However, the regression coefficient from our multivariate regression model with T (Eq. (2)) is positive (Table 4), indicating higher O_3 concentrations occur with more humid soils (note that in the model including [OVOCs] instead of T the soil H_2O term becomes non-significant). This could indicate changes in the plant emissions driven by soil moisture, promoting O_3 formation. Temperature can also affect the stomatal uptake of O_3 by changing the stomatal conductance of the plants. Stomatal conductance has been reported to increase with increasing T under constant humidity conditions (e.g. Urban et al., 2017). This could lead to an increase in O_3 uptake due to stomatal opening, but the real relationship is complicated. A review from Clifton et al. (2020a) concluded that the uptake of O_3 itself decreases the stomatal conductance due to changes in cell signalling pathways and turgor pressure. Further processing of O_3 inside the plant leaves can also cause e.g. cell death and stomatal sluggishness (Hoshika, 2015; Clifton et al., 2020b) thus affecting the uptake. Strong connection between the non-stomatal O_3 uptake and surface moisture has also been observed (Altimir et al., 2006). Lin et al. (2020), on the other hand, observed reduction in O_3 uptake by forests due to droughts caused by heatwaves. However, we do not have a variable in our model that would directly describe the O_3 uptake by vegetation. Soil H_2O does link to $[O_3]$ the effects of vegetation, however our results are not conclusive on the soil water effects as our model did not consider the potential time lag in the effects of soil moisture on plant emissions (Rissanen et al., 2020) and drawing solid conclusion regarding if part of the increase in $[O_3]$ (due to increase in air T) could be caused by changes in O_3 uptake by vegetation (caused by air T and/or related soil H_2O effect affecting plants) is not possible within our methodology.

4. Conclusions

In this study, we investigated the factors contributing to the apparent temperature dependency of O_3 concentration observed in Hyytiälä, Finland (estimated up to 2.65 ppb $^\circ C^{-1}$ with linear regression when considering the errors in both T and O_3 observations). Without considering the OVOCs, the temperature dependence of $[O_3]$ obtained from the mixed effects model is decreased to 0.87 ppb $^\circ C^{-1}$ when considering the basic components involved in O_3 formation and destruction, weather conditions, anthropogenic influence, and possible long-range transport of O_3. With these parameters, relatively good prediction accuracy (Pearson's r 0.84) for O_3 concentration is achieved for this dataset containing summer time daytime data. Note that this is achieved with a rather simple mixed effects regression model with no need to consider the detailed processes involved in O_3 formation and removal. Additional analysis with OVOCs included in the mixed effects model indicated that portion of the temperature dependence may be linked to OVOCs. While our methodology can identify OVOCs as potentially important factors for the observed O_3 concentration at this location, it cannot resolve the exact mechanism behind this relationship. OVOCs can act as indirect O_3 precursors (via oxidation with OH), but they can also be the products of BVOC oxidation which can be a loss process for O_3 (via ozonolysis) or a production process (via OH oxidation). To clarify which of the suggested pathways is truly responsible for the observed linkage, dedicated mechanistic studies are needed with in-depth chemical box models.

The mixed effects model deployed in this study identified factors that affect the observed O_3 concentrations and that were masked as the apparent temperature dependence when applying a simple linear regression. The factors to be included were based on their expected importance of O_3 concentration and availability of their long-term measured or calculated data. It should be noted, however, that our analysis was based on summertime (June–August) daytime data and, therefore, the importance of the variables should not be generalized for this site year-round. Our results indicate that the OVOC concentrations play an important role for variations of O_3 concentration and more investigations are needed to understand the detailed processes linking OVOC sources, sinks, and reactivities to observed O_3 concentrations, especially in boreal forest environments. The conditions at SMEAR II seem quite unique in comparison with the few other existing studies about apparent T dependence of O_3 concentrations, as shortly discussed in Section 3.1. Applying the mixed effects model approach to investigate the apparent and direct T dependence observed in these other environments may reveal temperature dependent processes involved in the O_3 formation/removal, with relatively small computational effort, which are hidden or not yet known. Tropospheric O_3 concentrations are observed to increase with temperature in both urban and rural locations around the world. As O_3 is also a harmful pollutant for both humans and vegetation, climate change driven increases in temperature could lead to unfavorable effects. Thus, investigation of the factors affecting the temperature dependence is essential. Compared to computationally expensive box models with detailed chemistry, applying a mixed effects model is a fast and relatively simple approach with no need to explicitly calculate all processes involved in ozone formation.

We encourage researchers to use this methodology to see if similar factors can be found to affect temperature dependence of $[O_3]$ in their data sets. Currently the model is defined with hourly data and applying it on daily averaged data would be indicative. However, lower resolution data might contain some bias as large fraction of the total variation would be neglected. For the data set used in this study, addition of the weather and BVOCs groups had the largest effects on the obtained slope for T. Depending on the environment, effect of the other groups may increase. For instance, relative contribution of local O_3 formation and long-range transport is expected to affect the relative importance of

including basic chemistry variables and back-trajectories in the model. The model in its current form is assumed to work properly in areas with similar meteorological conditions, with similar level of anthropogenic pollution and similar VOC profile to the SMEAR II location. To be used in very different environments, the model coefficients need further validation. Furthermore, in a very different environment, additional variables may become relevant, and should be added. Future work could include analysis with the model defined here and with coefficients obtained from our data to see how sensitive the current model is for different environments. Especially it would be interesting to see how the model performs, for example, in regions with extremely high temperatures (e.g. Steiner et al., 2010). Fitting new coefficients for different environments may also provide information on differences in dominating factors in defining [O_3] between environments or on factors missing from the current model.

Author contribution

T.Y., S.M. and S.I. designed the work. S.I. had the lead role in data analysis with supporting contribution from M.L. T.N., I.Y., T.P. and M.K. collected and provided the data. S.I., S.M., M. L., A.B, S.S. and T.Y. interpreted the results. S.I. wrote the paper with supporting contributions from S.M., A. B., S.S and T.Y. All co-authors commented and edited the manuscript.

Funding information

We acknowledge the following projects: ACCC Flagship funded by the Academy of Finland (grant numbers 337549 and 337550), the Academy of Finland project funding (grant no., 334792, 307537 and 311932), the Academy of Finland competitive funding to strengthen university research profiles (PROFI) for the University of Eastern Finland (grant no. 325022), the Jane and Aatos Erkko Foundation, the European Research Council (ERC) project ATM-GTP Contract No. 742206. The work of S.I, was financially supported by the University of Eastern Finland Doctoral Program in Environmental Physics, Health and Biology.

Declaration of competing interest

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

Data availability

Data is freely available through the online portal as stated in the Data availability section in the manuscript

Acknowledgements

We thank Krista Luoma from providing the pre-processed eBC data measured in SMEAR II and technical and scientific staff in Hyytiälä station. We thank Annele Virtanen for discussions we had during the development of this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2022.119315.

References

Aalto, J., et al., 2015. Onset of photosynthesis in spring speeds up monoterpene synthesis and leads to emission bursts. Plant Cell Environ. 38, 2299–2312.
Aaltonen, H., et al., 2013. Continuous VOC flux measurements on boreal forest floor. Plant Soil 369, 241–256.
Altimir, N., et al., 2006. Foliage surface ozone deposition: a role for surface moisture? Biogeosciences 3, 209–228.
Anttila, P., 2020. Air Quality Trends in Finland, 1994-2018. University of Helsinki.
Austin, E., et al., 2015. Ozone trends and their relationship to characteristic weather patterns. J. Expo. Sci. Environ. Epidemiol. 25, 532–542.
Baranizadeh, E., et al., 2014. The effect of cloudiness on new-particle formation: investigation of radiation levels. Boreal Environ. Res. 19, 343–354.
Bates, D., et al., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Software 1 (1), 2015.
Bloomer, B.J., et al., 2009. Observed relationships of ozone air pollution with temperature and emissions. Geophys. Res. Lett. 36.
Burzykowski, T., et al., 2013. Linear Mixed-Effects Models Using R : A Step-by-step Approach. Springer, New York, NY, UNITED STATES. New York.
Chambers, J.M., 1992. Chapter 4 in Statistical Models in S. Wadsworth & Brooks/Cole.
Clifton, O.E., et al., 2020a. Dry deposition of ozone over land: processes, measurement, and modeling. Rev. Geophys. 58, e2019RG000670.
Clifton, O.E., et al., 2020b. Stomatal conductance influences interannual variability and long-term changes in regional cumulative plant uptake of ozone. Environ. Res. Lett. 15, 114059.
Coates, J., et al., 2016. The influence of temperature on ozone production under varying NOx conditions – a modelling study. Atmos. Chem. Phys. 16, 11601–11615.
Cooper, O.R., et al., 2014. Global distribution and trends of tropospheric ozone: an observation-based review. Elementa: Science of the Anthropocene 2.
Dada, L., et al., 2017. Long-term analysis of clear-sky new particle formation events and nonevents in Hyytiälä. Atmos. Chem. Phys. 17, 6227–6241.
Demidenko, E., 2013. Mixed Models Theory and Applications with R. Wiley, Hoboken.
Deming, W.E., 1943. Statistical Adjustment of Data. John Wiley & Sons, New York.
Efron, B., et al., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 1, 54–75, 22.
Fao, 2020. Global Forest Resources Assessment 2020: Main Report (Rome).
Filella, I., et al., 2009. Short-chained oxygenated VOC emissions in Pinus halepensis in response to changes in water availability. Acta Physiol. Plant. 31, 311–318.
Freund, R.J., et al., 2006. Regression Analysis : Statistical Modeling of a Response Variable. Elsevier Academic Press, AMsterdam.
Fulgham, S.R., et al., 2020. Surface wetness as an unexpected control on forest exchange of volatile organic acids. Geophys. Res. Lett. 47.
Guenther, A., et al., 1995. A global-model of natural volatile organic-compound emissions. J. Geophys. Res. Atmos. 100, 8873–8892.
Hari, P., et al., 2005. Station for measuring ecosystem-atmosphere relations (SMEAR II). Boreal Environ. Res. 10, 315–322.
Harley, P., et al., 2007. Environmental controls over methanol emission from leaves. Biogeosciences 4, 1083–1099.
Hellen, H., et al., 2018. Long-term measurements of volatile organic compounds highlight the importance of sesquiterpenes for the atmospheric chemistry of a boreal forest. Atmos. Chem. Phys. 18, 13839–13863.
Hoshika, Yasutomo, et al., 2015. Ozone-induced stomatal sluggishness changes carbon and water balance of temperate deciduous forests. Sci. Rep. 5 (1) https://doi.org/10.1038/srep09871.
Ito, A., et al., 2009. Global chemical transport model study of ozone response to changes in chemical kinetics and biogenic volatile organic compounds emissions due to increasing temperatures: sensitivities to isoprene nitrate chemistry and grid resolution. J. Geophys. Res. Atmos. 114.
Junninen, H., et al., 2009. Smart-SMEAR: online data exploration and visualization tool for SMEAR stations. Boreal Environ. Res. 14, 447–457.
Kaufman, L., et al., 1990. Finding Groups in Data : an Introduction to Cluster Analysis. Wiley, New York.
Klingberg, J., et al., 2011. Ozone risk for vegetation in the future climate of Europe based on stomatal ozone uptake calculations. Tellus Dyn. Meteorol. Oceanogr. 63, 174–187.
Laban, T.L., et al., 2020. Statistical analysis of factors driving surface ozone variability over continental South Africa. J. Integr. Environ. Sci. 17, 1–28.
Li, H., et al., 2021a. Atmospheric organic vapors in two European pine forests measured by a Vocus PTR-TOF: insights into monoterpene and sesquiterpene oxidation processes. Atmos. Chem. Phys. 21, 4123–4147.
Li, K., et al., 2021b. Ozone Pollution in the North China Plain Spreading into the Late-Winter Haze Season, vol. 118. Proceedings of the National Academy of Sciences of the United States of America.
Liao, L., et al., 2011. Monoterpene pollution episodes in a forest environment: indication of anthropogenic origin and association with aerosol particles. Boreal Environ. Res. 16, 288–303.
Lin, M., et al., 2020. Vegetation feedbacks during drought exacerbate ozone air pollution extremes in Europe. Nat. Clim. Change 10, 444–451.
Loreto, F., et al., 2001. Ozone quenching properties of isoprene and its antioxidant role in leaves. Plant Physiol. 126, 993–1000.
Mcculloch, C.E., et al., 2008. Generalized, Linear, and Mixed Models. John Wiley & Sons, Hoboken, New Jersey, USA.
Mehtätalo, L., et al., 2020. Biometry for Forestry and Environmental Data : with Examples in R. CRC Press LLC, Boca Raton, FL.
Mellouki, A., et al., 2015. Atmospheric chemistry of oxygenated volatile organic compounds: impacts on air quality and climate. Chem. Rev. 115, 3984–4014.
Mikkonen, S., et al., 2020. Decennial time trends and diurnal patterns of particle number concentrations in a central European city between 2008 and 2018. Atmos. Chem. Phys. 20, 12247–12263.
Mikkonen, S., et al., 2019. Technical note: effects of uncertainties and number of data points on line fitting – a case study on new particle formation. Atmos. Chem. Phys. 19, 12531–12543.

Monks, P.S., 2005. Gas-phase radical chemistry in the troposphere. Chem. Soc. Rev. 34, 376–395.
Monks, P.S., et al., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. Atmos. Chem. Phys. 15, 8889–8973.
Mäki, M., et al., 2019. Boreal forest soil is a significant and diverse source of volatile organic compounds. Plant Soil 441, 89–110.
Ooka, R., et al., 2011. Influence of meteorological conditions on summer ozone levels in the central Kanto area of Japan. Procedia Environ. Sci. 4, 138–150.
Patokoski, J., et al., 2014. Winter to spring transition and diurnal variation of VOCs in Finland at an urban background site and a rural site. Boreal Environ. Res. 19, 79–103.
Patokoski, J., et al., 2015. Sources of long-lived atmospheric VOCs at the rural boreal forest site, SMEAR II. Atmos. Chem. Phys. 15, 13413–13432.
Pusede, S.E., et al., 2014. On the temperature dependence of organic reactivity, nitrogen oxides, ozone production, and the impact of emission controls in San Joaquin Valley, California. Atmos. Chem. Phys. 14, 3373–3395.
Pusede, S.E., et al., 2015. Temperature and recent trends in the chemistry of continental surface ozone. Chem. Rev. 115, 3898–3918.
R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rantala, P., et al., 2015. Annual cycle of volatile organic compound exchange between a boreal pine forest and the atmosphere. Biogeosciences 12, 5753–5770.
Rasmussen, D.J., et al., 2013. The ozone–climate penalty: past, present, and future. Environ. Sci. Technol. 47, 14258–14266.
Rencher, A., et al., 2012. Methods of Multivariate Analysis. Wiley.
Rickard, A., et al., 2021. The master chemical mechanism version MCM v3.2 [Online]. Available: http://mcm.leeds.ac.uk/MCMv3.2/. (Accessed 18 February 2021). Accessed.
Rissanen, K., et al., 2020. Stem emissions of monoterpenes, acetaldehyde and methanol from Scots pine (Pinus sylvestris L.) affected by tree–water relations and cambial growth. Plant Cell Environ. 43, 1751–1765.
Romer, P.S., et al., 2018. Effects of temperature-dependent NOx emissions on continental ozone production. Atmos. Chem. Phys. 18, 2601–2614.
Saunier, A., et al., 2017. Chronic drought decreases anabolic and catabolic BVOC emissions of quercus pubescens in a mediterranean forest. Front. Plant Sci. 8.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464, 4.
Seinfeld, J.H., et al., 2016. Atmospheric Chemistry and Physics: from Air Pollution to Climate Change. John Wiley & Sons.
Sillman, S., et al., 1995. Impact of temperature on oxidant photochemistry in urban, polluted rural and remote environments. J. Geophys. Res. Atmos. 100, 11497–11508.
Sousa, S.I.V., et al., 2013. Health effects of ozone focusing on childhood asthma: what is now known – a review from an epidemiological point of view. Chemosphere 90, 2051–2058.
Statistics Finland, 2019. *Official Statistics of Finland (OSF): Population structure* [online]. Helsinki: statistics Finland. Available: http://www.stat.fi/til/vaerak/index_en.html.
Stein, A.F., et al., 2015. Noaa's hysplit atmospheric transport and dispersion modeling system. Bull. Am. Meteorol. Soc. 96, 2059–2077.
Steiner, A.L., et al., 2010. Observed suppression of ozone formation at extremely high temperatures due to chemical and biophysical feedbacks. Proc. Natl. Acad. Sci. U. S. A 107, 19685–19690.
Urban, J., et al., 2017. Stomatal conductance increases with rising temperature. Plant Signal. Behav. 12, e1356534.
Vermeuel, M.P., et al., 2021. Simultaneous measurements of O3 and HCOOH vertical fluxes indicate rapid in-canopy terpene chemistry enhances O3 removal over mixed temperate forests. Geophys. Res. Lett. 48, e2020GL090996.
Wilks, S.S., 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. 9, 60–62, 3.
Wolfe, G.M., et al., 2011. Forest-atmosphere exchange of ozone: sensitivity to very reactive biogenic VOC emissions and implications for in-canopy photochemistry. Atmos. Chem. Phys. 11, 7875–7891.
Yan, Y., et al., 2018. Analysis of European ozone trends in the period 1995–2014. Atmos. Chem. Phys. 18, 5589–5605.
Yli-Juuti, T., et al., 2021. Significance of the organic aerosol driven climate feedback in the boreal area. Nat. Commun. 12, 5637.
Zhou, P., et al., 2017. Simulating ozone dry deposition at a boreal forest with a multi-layer canopy deposition model. Atmos. Chem. Phys. 17, 1361–1379.
Zuur, A., et al., 2009. Mixed Effects Models and Extensions in Ecology with R. Springer, New York, NY, UNITED STATES (New York).