Towards understanding the characteristics of new particle formation in the Eastern Mediterranean

Rima Baalbaki1, Michael Pikridas2, Tuja Jokinen1, Tiia Laurila1, Lubna Dada1, Spyros Bezantakos2, Lauri Ahonen1, Kimmo Neitola3,4, Anne Maisser3, Elie Bimenyimana3, Aliki Christodoulou1,2, Florin Ungu2, Chrysanthos Savvides5, Katrianne Lehtipalo1,6, Juha Kangaslouma1, George Biskos2, Tuukka Petäjä1, Veli-Matti Kerminen1, Jean Sciare2, Markku Kulmala1

1Institute for Atmospheric and Earth System Research (INAR) / Physics, Faculty of Science, University of Helsinki, P.O. Box 64, Helsinki, 00014, Finland
2Climate & Atmosphere Research Centre (CARE-C), The Cyprus Institute, P.O. Box 27456, Nicosia, CY-1645, Cyprus
3IMT Lille Douai, Université de Lille, SAGE - Département Sciences de L’Atmosphère et Génie de L’Environnement, 59000, Lille, France
4Ministry of Labour, Welfare and Social Insurance, Department of Labour Inspection (DLI), Nicosia, Cyprus
5Finnish Meteorological Institute, Helsinki, Finland

Correspondence to: rima.baalbaki@helsinki.fi

Abstract

To quantify the contribution of new particle formation (NPF) to ultrafine particle number and CCN budgets, one has to understand the mechanisms that govern NPF in different environments and its temporal extent. Here, we study NPF in Cyprus, an Eastern Mediterranean country located at the crossroads of three continents, and affected by diverse air masses originating from continental, maritime, and desert-dust source areas. We performed one-year continuous measurements of aerosol particles down to ~1 nm in diameter, for the first time in the Eastern Mediterranean and Middle East (EMME) region. These measurements were complemented with trace gas data, meteorological variables, and retroplume analysis. We show that NPF is a very frequent phenomenon at this site and has higher frequencies of occurrence during spring and autumn. NPF events were both of local and regional origin, and the local events occurred frequently during the month with the lowest NPF frequency. Some NPF events exhibited multiple onsets, while others exhibited particle apparent shrinkage in size. Additionally, NPF events were observed during the night-time and during episodes of high desert dust loadings. Particle formation rates and growth rates were comparable to those in urban environments, although our site is a rural one. Meteorological variables and trace gases played a role in explaining the intra-monthly variability of NPF events, but did not explain why summer month had the least NPF frequency. Similarly, pre-existing aerosol loading did not explain the observed seasonality. The month with the least NPF frequency were associated with higher H2SO4 concentrations but lower NOx concentration, which is an indicator of anthropogenic influence. Air masses arriving from the Middle East were not observed during these months, which could suggest that precursor vapors important for nucleation and growth are transported to our site from the Middle East. Further comprehensive measurements of precursor vapors are required to prove this hypothesis.

1 Introduction

Atmospheric new particle formation (NPF) is the process by which oxidized precursor gases initially form molecular clusters that then further grow in size by multi-component condensation (Kulmala et al., 2014). A multitude of research studies have focused on this phenomenon over the past two decades because it is a large...
source of the global aerosol particle number and cloud condensation nuclei (CCN) load (Gordon et al., 2017; Merikanto et al., 2009; Pierce and Adams, 2009; Wang and Penner, 2009; Yu and Luo, 2009; Kerminen et al., 2012; Spracklen et al., 2006; Spracklen et al., 2008). Owing to the complex nature and non-linearity of atmospheric processes, studies on NPF in the literature include atmospheric observations (e.g. Kulmala et al., 2013; Ehrl et al., 2014; Bianchi et al., 2016; Yao et al., 2018; Williamson et al., 2019; Baccarini et al., 2020; Dall’Osto et al., 2018), chamber experiments (e.g. Sipila et al., 2010; Tröstl et al., 2016; Wang et al., 2020; Lehtipalo et al., 2016; Kirkby et al., 2011), and theoretical computational studies (e.g. Kurten et al., 2008; Riipinen et al., 2011; Olenius and Riipinen, 2017). The collective scientific outcome from these studies is essential to understand the mechanisms and characteristics of NPF (Kerminen et al., 2018; Lee et al., 2019; Chu et al., 2019) and how it affects the global climate (e.g. Spracklen et al., 2006; Gordon et al., 2017).

The frequency, strength, and spatio-temporal extent of NPF are mainly governed by three factors: the prevailing meteorological conditions, the availability of gaseous precursors, and the pre-existing concentrations of aerosol particles (Kerminen et al., 2018; Lee et al., 2019; Nieminen et al., 2018). These atmospheric conditions differ in space and time. Atmospheric conditions are distinct over the Mediterranean basin, especially over the Eastern Mediterranean and Middle East (EMME). This region has been identified as a hotspot for atmospheric and climate change research (Lelieveld et al., 2016; Giorgi and Lionello, 2008). It is surrounded by three continents and is affected by continental, maritime, and desert-dust pollution sources (Lelieveld et al., 2002). The surrounding complex orography of the Mediterranean affects atmospheric dynamics and boundary layer processes on different scales (Kostopoulou and Jones, 2007b, a). Further, the dry and hot weather throughout most of the year, with strongly increasing heat extremes, enables intense photochemistry (Lelieveld et al., 2016).

NPF studies over the Mediterranean have focused on the north western basin (Pétaü et al., 2007; Cusack et al., 2013; Berland et al., 2017; Carnerero et al., 2018; Rose et al., 2015; Brines et al., 2015; Hamed et al., 2007; Laaksonen et al., 2005; Casquero-Vera et al., 2020), whereas NPF studies in the eastern basin have been conducted mainly in Greece (Pétaü et al., 2007; Berland et al., 2017; Kalivitis et al., 2015; Kalivitis et al., 2019; Pikridas et al., 2012; Kalkavouras et al., 2019; Kalkavouras et al., 2020; Kopanakis et al., 2013; Siakavaras et al., 2016; Kalkavouras et al., 2017) and very recently in Cyprus (Brilke et al., 2020; Debevec et al., 2018) and Jordan (Hussein et al., 2020). These studies include both short-term campaigns and long-term observations.

Based on long-term measurements, the annual frequency of NPF over the Mediterranean varies between 10 and 36% (Hussein et al., 2020; Kalivitis et al., 2019; Kalkavouras et al., 2020; Kopanakis et al., 2013). The seasonal cycle has a typical maximum during spring (Kalkavouras et al., 2020; Kopanakis et al., 2013; Kalivitis et al., 2019; Pikridas et al., 2012), even though in some urban background sites the highest frequency was observed during summer (Hussein et al., 2020; Hamed et al., 2007). NPF was associated with a high increase in nucleation mode particles in most of the studies. For instance, Carnerero et al. (2018) showed that the impact of NPF on ultrafine particles is much higher than that of traffic near the highly polluted city center of Madrid.

The condensation sink, which is a measure of the pre-existing aerosol surface area, was reported to be lower during NPF events in the Po valley, Corsica, and Crete (Hamed et al., 2007; Berland et al., 2017; Pikridas et al., 2012), while NPF proceeded under both clean and polluted conditions in Barcelona (Cusack et al., 2013), Marseille and Athens (Pétaü et al., 2007). The effect of meteorological conditions on NPF occurrence varied among studies. Simultaneous NPF events were observed in several stations, illustrating that the spatial extent of NPF events can vary from tens of kilometers (Carnerero et al., 2018) to several hundred kilometers (Kalkavouras et al., 2017; Kalkavouras et al., 2020; Berland et al., 2017; Rose et al., 2015; Casquero-Vera et al., 2020). In the Po valley, the production of CCN from NPF was estimated to be comparable to that originating from primary sources (Laaksonen et al., 2005). Similarly, NPF was associated with a strong increase in CCN concentrations in Finokalia and Santorini (Kalkavouras et al., 2019; Kalkavouras et al., 2017; Kalivitis et al., 2019; Chu et al., 2019) and it affects the global climate (e.g. Spracklen et al., 2006; Gordon et al., 2017).
However, the impact of the increased CCN concentrations on cloud droplet number was shown to be limited by water availability (Kalkavouras et al., 2017). In Cyprus, mainly in Paphos, Gong et al. (2019) observed several NPF events where newly-formed particles grew into the CCN size range, with NPF events being observed on 9 out of 27 measurement days during April 2017 (Brilke et al., 2020). At a more inland site, NPF was observed on 14 out of 20 days of measurements on March 2015 (Debevec et al., 2018). Since these studies were less than a month long, further comprehensive measurements are required to unveil the role of NPF in the atmospheric processes taking place in the EMME region.

The aim of this study is to characterize the seasonal cycle of new particle formation events in the less represented area of the EMME region. Our measurements were conducted at a rural background site on the island of Cyprus which lies at the crossroads of three continents in the Eastern Mediterranean. We report the first long-term analysis of particle number size distribution in the area, down to sizes where the initial formation occurs. We further explore the role of sulfuric acid, which is one of the key gas phase precursor for cluster formation, and other atmospheric variables in initiating NPF at this site.
2 Materials and Methods

2.1 Measurement site

Cyprus is an island country in the Eastern Mediterranean. It is the third most populous island in the Mediterranean Sea, and the third largest in size with an area of 9251 km² (Figure 1). The measurements reported in this work were carried out at the Agia Marina Xyliatos station of the Cyprus Atmospheric Observatory (CAO-AMX; Sciare, 2016), which is a rural background station that operates under the co-operative programme for monitoring and evaluation of the long-range transmission of air pollutants in Europe (EMEP) and the European Research Infrastructure for the observation of Aerosol, Clouds and Trace Gases (ACTRIS) networks, while at the same time it is a designated regional Global Atmospheric Watch (GAW) station. The station (35.038692° N, 33.057850 ° E) is located close to the villages of Agia Marina (~630 inhabitants) and Xyliatos (~150 inhabitants) and has an elevation of 532 m above sea level. The proximity of the site is surrounded by vegetation mainly oak and pine trees and Maquis shrubland as it lies at the northern-eastern foothills of the Troodos Mountains. Agriculture areas surround the site from the north direction and are approximately 4 km away. The nearest main urban agglomeration is at least 35 km away. Therefore, it is not directly affected by any major local pollution source, excluding some limited traffic to the nearby Forestry Department premises.

The variable air mass origins at CAO-AMX from three different continents allow a representative description of NPF processes for the EMME region as a whole. The weather at CAO-AMX is characterized by hot dry summers and mild, rainy winters. The daily mean temperature is ~19°C and ranges between 1 and 36°C, the daily mean relative humidity is ~55% and ranges between 13 and 82%, and the daily mean ozone level is ~48 ppb ranging between 26 and 77 ppb (Kleanthous et al., 2014). The most common (> 65% occurrence) wind pattern reaching the site is the northerly “Etesian” winds transporting pollutants from both Europe and Turkey, but more frequently from mainland of Turkey (Pikridas et al., 2018). The remaining air masses originate from North Africa, the Middle East and westerlies air masses that spend several days above the sea before reaching Cyprus. The variable air mass origins at CAO-AMX are characterized by typical Mediterranean weather conditions and have an area of 9251 km². The island of Cyprus hosts a large number of inhabitants (approximately 1 million) and Xyliatos (~150 inhabitants) and is part of the third largest agglomeration in Cyprus (ACTRIS; Sciare, 2016).

AMX is a rural background station that operates under the co-operative programme for monitoring and evaluation of the long-range transmission of air pollutants in Europe (EMEP) and the European Research Infrastructure for the observation of Aerosol, Clouds and Trace Gases (ACTRIS) networks, while at the same time it is a designated regional Global Atmospheric Watch (GAW) station. The station (35.038692° N, 33.057850 ° E) is located close to the villages of Agia Marina (~630 inhabitants) and Xyliatos (~150 inhabitants) and has an elevation of 532 m above sea level. The proximity of the site is surrounded by vegetation mainly oak and pine trees and Maquis shrubland as it lies at the northern-eastern foothills of the Troodos Mountains. Agriculture areas surround the site from the north direction and are approximately 4 km away. The nearest main urban agglomeration is at least 35 km away. Therefore, it is not directly affected by any major local pollution source, excluding some limited traffic to the nearby Forestry Department premises.

The A11 nCNC is composed of a Particle Size Magnifier (PSM; Airmodus A10) and a Condensation Particle Counter (CPC; Airmodus A20). The overall length of the inlet sampling tube was 60 cm. The PSM was operated in a scanning saturator flow mode between 0.1 and 1.3 liter per minute (lpm), corresponding to a cutoff diameter range of approximately 1.1 to 2.5 nm. It was equipped with an inlet system that performs background (zero) measurements three times a day at random time intervals, and a core sampling piece for minimizing line losses of sub-3 nm particles (Figure S1). The duration of the background measurements was set to 12 minutes, which is equivalent to 3 full size scans. From June 2018 onwards, the nCNC was additionally equipped with a diluter to reduce the humidity of the sampled air. This procedure was necessary because the
water content of the air at the measurement site was too high. The water present in the sample air was mixed with butanol inside the CPC of the nCNC and rendered it measuring zeros. Further information about the dilution design, its operation and effect on the data can be found in the supplementary information (SI) Sect. 2.2.

The NAIS is a mobility spectrometer designed to determine the number size distribution of ions in the mobility diameter range of 0.8 – 42 nm, as well as total (naturally charged and neutral) aerosol particles in the mobility diameter range of ~2 – 42 nm. The instrument operates at the flow rate of ~54 lpm. The length of the NAIS sampling tube was 65 cm, with an inner diameter of 30 mm.

The SMPS used in this study was composed of a TSI 3081 long Differential Mobility Analyzer (DMA) and a TSI 3025a CPC. It was operated to measure the aerosol particle size distribution between 15 and 740 nm. The aerosol and sheath flow were checked weekly and were set to 0.3 and 3 lpm, respectively. The SMPS was sampling using an 80-cm long vertical inlet. Drying was achieved using a short nanoflon dryer, and charge neutralization was achieved by a GRIMM 5522-A, Americium-241, bipolar neutralizer.

### 2.2.2 Ancillary measurements

Complementary meteorological data (temperature, relative humidity, solar radiation, rainfall, pressure, wind speed, and wind direction) were measured with a time resolution of five minutes at an elevation of 10 m from the ground in the nearby village of Xyliatos (35.0140917 N, 33.0492028 E, located 2.85 km from the measurement site). Air pollutants (ozone, carbon monoxide, nitrogen oxides, sulfur oxide, PM$_{10}$, and PM$_{2.5}$) were measured at the collocated EMEP station ~20 m from the main measurement container, and these data had a time resolution of one hour. Additional details about the set-ups and the instrument used can be found in Kleanthous et al. (2014) and Pikridas et al. (2018).

### 2.3 Data handling

*nCNC*: The scanning nCNC data were inverted into a size distribution with the Kernel inversion method presented by Lehtipalo et al. (2014), but using customized kernels which follow the instrument specific detection efficiency calibration curves. The following diameters were used in the inversion: 1.1 nm, 1.3 nm, 1.5 nm, and 2.4 nm. The choice of the inversion method was made after a comprehensive comparison between the Kernel method and the Expectation and Minimization (EM) method (Cai et al., 2018; Chan et al., 2020). Additional details about the comparability of the two methods and the utilized inversion parameters are presented in Sect. 2.3 of the SI. After inversion, the data were further corrected for line losses using the method suggested by Fu et al. (2019) for the sampling line downstream of the core sampling inlet, and using the Gormley and Kennedy equation for the line losses inside the 6-cm-long core sampling piece (Gormley and Kennedy, 1948).

*NAIS*: The NAIS data were inverted with the instrument specific algorithm (done by the NAIS SPECTOPS software). The data were later corrected for line losses using the Gormley and Kennedy equation for laminar flow (Gormley and Kennedy, 1948). It is essential to note that the flow through the sampling inlet of the NAIS actually lies in the transient regime (Re = 2376), however the penetration efficiency using this inlet was comparable for laminar flow and turbulent flow (calculated using the equation of turbulent inertial deposition from Brockmann (2011)), thus we used the correction based on laminar flow (Figure S4).

*SMPS*: The data from the SMPS were inverted using the TSI’s Aerosol Instrument Manager (AIM, version 9.0) software. Afterwards, line loss correction was applied using Gormley and Kennedy equation. Additional corrections based on lab calibrations were also applied to account for the CPC detection efficiency curves.
**Full Particle Size Distribution (PSD):** The data from the three particle sizing instruments were used to
reconstruct the full particle size distribution with a temporal resolution of 5 minutes between 1.1 and 736 nm
(nCNC: 1.1 to 2.4 nm; NAIS particle mode: 2.4 to 30 nm; SMPS: 30 to 736 nm). However, the SMPS measured
dry aerosol particle number distributions, which can differ from the ambient aerosol particle number size
distribution. Thus, we back-calculated the distribution of the SMPS at ambient conditions from the dry
distribution using the hygroscopicity model of Petters and Kreidenweis (2007) and mean kappa values.
Additional information about these calculations and its effect on sink calculations are presented in Sect. four
of the SI material. The SMPS distribution at ambient conditions was reconstructed up to 1500 nm. This does
not imply that the measurement range was extended to 1500 nm, but rather that now we account for particles
that were originally of sizes up to 1500 nm but were dried to sizes below 736 nm in the SMPS sampling line.
Additionally, since the NAIS is known to overestimate concentrations in the particle mode, the overlapping
measurement range with the SMPS was used to further correct the NAIS data assuming that the NAIS
overestimate concentrations uniformly over the whole measurement range, which is a reasonable assumption
for old NAIS models based on calibration results (Gagné et al., 2011; Kangasluoma et al., 2020). Finally, the
PSD data was run through a 2D median filtering algorithm with a 3x3 neighborhood window. Moreover,
the data was manually checked for the success of the outlier and noise removal techniques.

**Complementary data:** Gas and meteorology data sets were run through an outlier removal algorithm and
filtered for erroneous samples. The outlier detection method was based on removing data points that are more
than three standard deviations from a moving median (Davies and Gather, 1993; Pearson et al., 2016).

### 2.4 Event classification

The reconstructed full particle size distribution daily plots were used to categorize measurement days into NPF
event days, non-event days, and undefined days based on a classification that combines the schemes reported
in literature (Dal Maso et al., 2005; Hirsiiko et al., 2007; Manninen et al., 2010; Kulmala et al., 2012). The
classification of events utilizing PSD data that extends below 10 nm, which is a typical measurement limit for
most SMPS systems, improves the event classification and allows better identification of event days that would
otherwise be classified as undefined or non-events if only PSDs above 10 nm were used (Leino et al.,
2016; Dada et al., 2018; Brilke et al., 2020). In addition, spectra of total particles (both neutral and charged) are
usually easier to visually classify than those corresponding to charged particles (measured by the ion mode of
NAIS) because atmospheric nucleation is dominated by neutral processes (Kontkanen et al., 2013; Kulmala et
al., 2013; Wagner et al., 2017). In addition, the concentration of the growing mode in the charged spectra is
lower for the smaller particle sizes, and increases with diameter as the probability of cluster ions attaching to
the growing neutral particles increases (Gonsor et al., 2014). Thus, it could be visually difficult to determine
if particle nucleation starts from the smallest sizes when looking at charged spectra only. In contrast, one
should not neglect looking at charged spectra because it might show sign preference or ion induced nucleation
events (Rose et al., 2018).

### 2.5 NPF specific parameters

**Condensation sink (CS):** is a loss term for condensable vapors used to describe their loss rate by condensation
onto pre-existing aerosol surface. This term was first introduced by Kulmala et al. (2001) and it is derived based
on condensing vapor mass flux to the particles in the continuum regime and applying the transitional correction
factor ($\beta_m$) proposed by Fuchs and Sutugin (1971):

$$
CS = 4\pi D \sum_i \beta_m N_i = 2\pi D \sum_i \beta_m d_i N_i,
$$

(1)
where \( r, d_p \) and \( N \) are the particle radius, diameter and number concentration, respectively, in the size class \( i \), and \( D \) is the diffusion coefficient of the condensing vapor calculated as recommended by Fuller et al. (1966):

\[
D(H_2SO_4, air) = \frac{0.001T^{1.75}}{P} \frac{1}{\sqrt{M_{H_2SO_4} + \frac{1}{M_{air}}}} \frac{1}{(\sqrt{\nu_{H_2SO_4}} + \frac{1}{\sqrt{V_{air}}})^2},
\]

where \( T \) is the temperature, \( M \) is the molar mass, \( P \) is the atmospheric pressure, and \( V \) is the diffusion volume.

Here, \( CS \) was calculated assuming that sulfuric acid is the main condensing vapor.

**Coagulation sink (CoagS)** is a loss term for freshly formed particles used to describe their loss rate by Brownian coagulation to pre-existing aerosol surface (Kulmala et al., 2001). It is calculated as:

\[
CoagS(d_p) = \sum K_{ij}N_j,
\]

where \( K_i \) is the Fuchs form of the Brownian coagulation coefficient (Fuchs, 1964; Seinfeld and Pandis, 2012).

**Apparent growth rate (GR)** is the rate of change in the diameter, \( d_p \) that represents the growing particle population. It was calculated here using the NAIS data for negatively charged ions, positively charged ions, and total particles (charged + neutral) by the appearance time method (Lehtipalo et al., 2014). First, the time to reach 50% of the maximum concentration is determined and then the growth rate is derived as the slope of the linear fit between the diameters and time:

\[
GR = \frac{d\Delta d_p}{dt}
\]

We calculated GR at three different size ranges: between 1.5 and 3 nm (GR\(_{1.5-3}\)), between 3 and 7 nm (GR\(_{3-7}\)) and between 7 and 20 nm (GR\(_{7-20}\)).

**Event start and end times** were determined based on the time evolution of the 2-4 nm particles which is the size range suggested by Dada et al. (2018). Using this size range, we are able to capture the early stages of the event which is unachievable if the measured PSD starts from bigger sizes. Thus, computed event start and end times might differ across studies depending on the instrument used. An event start is determined by an increase in the 2-4 nm particle concentration above the nighttime level which last for at least an hour. An event end time is determined when the 2-4 nm particle concentration decrease to background levels. In case of multiple events within a one-day window, the event start and end times were taken from the start of the first event until the end of the last event, respectively.

**Particle formation rate (J)** is the rate at which aerosol particles of certain size are formed in the atmosphere. It quantifies the intensity of the NPF events, and it is calculated by rearranging the equation describing the time evolution of the particle number concentration (Kulmala et al., 2012). \( D_p \) in this equation refers to the smaller limit of the size bin used in the calculation of the formation rate. We calculated \( J \) at three sizes: 1.5 nm (\( J_{1.5} \)), 3nm (\( J_3 \)) and 7nm (\( J_7 \)); the upper size limits used were 3, 7, and 20 nm, respectively. GR was calculated as the mean of the three GR measurements (negative ions, positive ions and total particles) and was considered constant within the event start and end times. Outside the event times and during non-events the GR term was considered equal to zero.
The first term in eq. 5 represents the time evolution of particle number concentration $N_{dp}$, the second term represents the coagulation losses due to larger aerosol particles, and the third term represents the condensational growth to sizes bigger than the considered size range.

### 2.6 Sulfuric acid proxy

Sulfuric acid is one of the key gas-phase compounds identified to contribute to new particle formation (e.g. Weber et al., 1996; Sipilä et al., 2010). As direct measurements of sulfuric acid is challenging, a suite of proxies for the sulfuric acid concentrations are derived that facilitate calculation of gas phase sulfuric acid from ancillary observations (Dada et al., 2020; Mikkonen et al., 2011; Petäjä et al., 2009; Lu et al., 2019; Weber et al., 1997). In this study, the sulfuric acid proxy was calculated using the new method by Dada et al. (2020) for a rural site, which was developed based on observations from the same site of this study:

$$\left[H_2SO_4\right]_{rural} = \frac{CS}{2 x (2 x 10^{-9})} + \left(\frac{CS}{2 x (2 x 10^{-9})}\right)^2 + \left[SO_2\right] \times \left(9 x 10^{-9} \times GlobRad\right)^{\frac{1}{2}}$$ (6)

This proxy does not only consider the formation of H$_2$SO$_4$ from SO$_2$ via OH oxidation and the loss of H$_2$SO$_4$ onto pre-existing particles (condensation sink), but it also includes loss of H$_2$SO$_4$ via atmospheric clustering starting from H$_2$SO$_4$ dimer formation.

### 2.7 Air mass origin analysis

Air mass origins for the entire measurement period were modeled using the Lagrangian particle dispersion model FLEXpart (FLEXible PARTICle dispersion model), version 8.23, in a backward mode (Stohl et al., 2005), with meteorological (0.5x0.5°, 6 h starting from midnight UTC) NCAR (ds 0.94) data as input. We used “species” 1 (tracer), which do not include wet or dry deposition and assumes an infinite lifetime for the particles, as the tracer released to model the retroplumes. Retroplumes replaces simple back trajectory calculations in the interpretation of atmospheric trace substance measurements, and were traced back in time for 5 days using CAO-AMX as the receptor site. Air masses were categorized to source regions based on the potential emission sensitivity (PES) for the lowest 1 km above ground level (agl), following the classification method of Pikridas et al. (2010). In general, a retroplume was attributed to a region in the case that this had a PES value above 0.9 ns kg$^{-1}$. The classification scheme of the source regions took into consideration dominant air mass paths shown by Pikridas et al. (2018) and the different sources of PM with characteristic chemical fingerprint. As a result the predominant northerly air masses were categorized into “Europe” and “NW Asia” (namely Turkey), assuming different emissions related to SO$_2$, “N. Africa” and “SW Asia” are both source areas of dust particles but with distinct emission levels, with the former being associated with more elevated concentrations. The “Asia sector” was distinguished to point out that air masses from this specific source region scarcely reach the receptor site, while the source region “Local” refers to stagnant conditions. Last, the “Marine” sector is associated with the lowest levels of ambient PM. In total, seven source regions were identified similar to the ones presented by Pikridas et al. (2018) except that in our analysis, the “West Turkey” sector was merged to the “NW Asia” sector.

### 2.8 Identification of days with high dust loading

Measurement dates with high dust loading were identified using the VI-PM1 online method proposed by Drinovec et al. (2020). This method couples a high-flow virtual impactor (VI) sampler, which concentrates coarse particles, with an aerosol absorption photometer. More details about the calculations and a list of the identified dust days can be found in Sect. 5 of the SI material.
3 Results and Discussion

In the course of identifying NPF events, the PSD spectrum is usually analyzed, mainly at sizes below 25 nm where one can detect the emergence of new aerosol particles, and then the particle growth to larger sizes is followed. Since little is known about particle number size distributions from the EMME region, we will first present the seasonal and diurnal variability of particle number concentration in different PSD modes (Sect. 3.1). Then, we will identify and characterize NPF events (Sect. 3.2). Following, we will quantify and analyze relevant parameters that describe NPF events (Sect. 3.3) and use those parameters, together with meteorological variables, to understand why and when NPF occurs at our site (Sect. 3.4). We further present a regression and a classification analysis in Sect. 3.5. All the data in this manuscript are presented in local time (UTC+3 from 25 March 2018 to 28 October 2018, and UTC+2 during the rest of the campaign). Unless otherwise indicated, we mainly focus on daytime data having global radiation > 50 W m⁻² because it is the time period relevant for most NPF events, but we also briefly mention some night-time events. For reference, the monthly range of day hours having global radiation > 50 W m⁻² is presented in Figure 5c.

3.1 Seasonal and diurnal variability of number concentration in different modes

Figure 2 presents the monthly percentiles boxplots (25th, 50th and 75th) and the mean averages of the cluster mode [-1–3 nm], nucleation mode [3–25 nm], Aitken mode [25–100 nm], and accumulation mode [100-1000 nm] particle number concentrations computed from daily data with global radiation > 50 W m⁻² (daytime conditions). A clear seasonal pattern is depicted which is distinct across the different particle modes. The cluster mode particles had two peaks, one in spring and another in autumn with a clear drop during the summer. The monthly box plots also show high variability in daily concentrations throughout most months, except August and September. This variability seems highest during February. The nucleation mode particles were also highest during the spring and lowest during the summer. The autumn concentrations did not exhibit another peak but were rather similar to the summer concentrations in terms of median values but they exhibited higher variability. The cluster and nucleation mode concentrations can be directly linked to the NPF activity, especially in sites where direct emissions of particles having these size ranges are minimal, which is the case for our site. While the high concentrations of cluster mode particles during spring was associated with high concentrations of nucleation mode particles, this did not hold for autumn, which might indicate that condensable vapors were not as available to grow the particles to nucleation size. The Aitken mode exhibited higher concentrations during the spring months followed by a decreasing pattern, which could either suggest more growth from NPF to Aitken sizes or higher emission/transportation of primary particles during spring. The accumulation mode had its maximum during the warm months, except during July which did not follow the pattern of other months. Previous long-term measurements of PM₉, at this site have a similar pattern with higher concentrations during the warm period of the year and minimum during winter (Pikridas et al., 2018). This maximum during the summer is mainly explained by the enhanced transport of polluted air masses from the north sector, combined with the lack of precipitation and overall dry conditions during Eastern Mediterranean summer (Pikridas et al., 2018). Last, it is worth mentioning that during February, the concentrations of particles in all modes did not follow the overall trend. It exhibited lower concentrations of cluster, nucleation, and Aitken mode particles and higher concentration of accumulation mode particles than the nearby month.

The diurnal variation (at radiation > 50 W m⁻²) of the cluster and nucleation mode particles exhibited a clear cycle, with the highest values recorded between 9:00 and 15:00 am and the maximum at 11:00 (Figure 3a & b). There was a slight time difference between the appearance of the cluster mode particles that of the nucleation mode particles which could only be seen in the 5-min data. The Aitken mode had a less distinct diurnal cycle having the peak at later hours of the day, which might indicate that these particles have possibly grown from the cluster and nucleation modes (Figure 3c). The accumulation mode, on the other hand, did not.
exhibit any clear diurnal cycle, which might suggest that these particles are not emitted or produced from any local source but are rather long-range transported. They can be aged primary particles, or particles originating from NPF taken place 1-3 days earlier in arriving air masses (Figure 3, d).

3.2 General character of the NPF events

New particle formation has been detected to occur in a variety of environments within the troposphere (Kerminen et al., 2018; Lee et al., 2019; Nieminen et al., 2018). Typically, the appearance of clusters is detected in the morning hours followed by subsequent growth. The occurrence of new particle formation events is determined by examining the time evolution of the aerosol number size distributions (e.g. Kulmala et al., 2012). Throughout the one-year measurement campaign (365 days), 207 (56.7 %) days were identified as event days, 119 (32.6 %) days were identified as non-event days, 31 (8.5 %) days were undefined days and 8 (2.2 %) days did not have data mainly due to power cuts at the station (Figure 4). The annual median NPF frequency at CAO-AMX calculated without accounting for days with no data amounts to 58 % which belongs to the high end of the global NPF frequency distribution (Nieminen et al., 2018) with the highest frequency being measured in South Africa (86 %; Hirikko et al., 2012). High frequency of NPF occurrence is also observed at Saudi Arabia (73 %; Hakala et al., 2019).

NPF took place throughout the year, but it had a clear seasonal pattern with a broad spring maximum, less pronounced autumn maximum, and slightly lower frequencies during other times of the year. The months with the highest NPF frequencies were March and April, while June and August had the lowest frequencies. This seasonal pattern of NPF frequency is very similar to that recorded at the Finokalia atmospheric observation station in Crete (Kalivitis et al., 2018), which is a nearby Eastern Mediterranean site having similar synoptic conditions. Nonetheless, monthly NPF frequency at Finokalia ranged between ~17 and 42 % which is substantially lower than the range reported here (33 - 86 %). The higher NPF frequency at CAO-AMX could partially be due to the use of PSD data that starts from the ~1 nm size range, which facilitates NPF classification especially during days when the particle growth does not pass the 10 nm size or does not continue for several hours. We compared the NPF classification using SMPS data only and that using full PSD for time periods when SMPS data were available, and attained 30 % less event days classified. Another factor that could contribute to the higher NPF frequency is the surrounding forest nature which emits VOCs that oxidize in the atmosphere and contribute to particle growth (Ripininen et al., 2011).

We further separated the NPF event days into class I or class II events, or into the so-called “bump” events (Manninen et al., 2010). The calendar of event classification is presented in Figure S7 and examples of event types are given in Figure S8. Class I events differ from class II events not by the strength of the event but rather by the ability to calculate the particle growth rate for such event, meaning that the growing mode diameter or concentration does not fluctuate strongly. Bump events are NPF events where a burst of nucleation mode particles is seen but the particles do not usually grow past the ~10 nm size, and the duration of these events is typically short. The calculation of growth rates for these events is sometimes problematic because the growth happens very fast (in less than 15 minutes) and it cannot be captured by the time resolution of the measuring instrument. In the literature, these events have been called “bursting events” (Dall’Osto et al., 2017), “hump events” (Vakkari et al., 2011; Yli-Juuti et al., 2009), “suppressed events” (Chen et al., 2017), “stationary NPF events” (Groß et al., 2018) or “weak NPF events” (Lee et al., 2020). The fraction of these events were highest during the month with the lowest NPF frequency (mainly during summer), which could imply that during these months less amount of condensing vapors was present to grow the particles to bigger sizes or extend the event duration (Figure 5).
The NPF events started almost always from the sub-3-nm range at CAO-AMX. The apparent growth reached a diameter of 20 nm on 25% of event days (Figure 5a), thus it could have been difficult to identify those days if we have relied solely on SMPS measurements which suffer from high losses and low counting statistics in the sub-10-nm size range (Briñol et al., 2020; Kangaslouama et al., 2020; Wiedensohler et al., 2012). Additionally, it was difficult to distinguish the growing mode at sizes above 50 nm mainly because of background aerosols and fluctuating air masses. This implies that particles growing from NPF might have been able to grow to bigger sizes, but their identification from the PSD spectrum was not possible. The median event duration was ~ 5.4 hours (Figure 6b). The events typically started two to four hours after sunrise and ended seven to eleven hours after sunrise (Figure 6c), similar to what was observed by Dada et al. (2018).

Another feature of NPF events observed at CAO-AMX was the occurrence of two or three consecutive daytime nucleation events (Figure 7). These multiple events occurred on ~20% of the recorded event days. Similar observations were reported in South Africa, and they were mainly attributed to changes in air masses, interruptions by clouds, and boundary layer dynamics and its relation to the amount of vapors present (Hirsikko et al., 2013). Salma and Németh (2019) have also showed that NPF events with broad or multiple onsets are abundant in the urban environment of Budapest, Hungary.

We also observed events with a decreasing mode diameter, sometimes referred to as shrinkage events. These events were mainly observed in the NAIS ion mode, while some of them were also observed in both ion and total particle spectrum (Figure 8). These types of events have been observed in multiple environments and are usually attributed to particle evaporation triggered by elevated temperatures or size-dependent dilution caused by wind- or boundary layer development-mixing, or changes in air masses bringing small particles to the measurement site (Alonso-Blanco et al., 2017; Backman et al., 2012; Cusack et al., 2013; Hakala et al., 2019; Kivekäs et al., 2016; Salma et al., 2016b; Skrabalova et al., 2015; Tsagkogeorgas et al., 2017; Yao et al., 2010; Young et al., 2013; Zhang et al., 2016; Carnerero et al., 2018).

We spotted a few events with nighttime clustering, which could reflect a chemistry that does not depend on photo-oxidation (Figure 9). These events occurred mainly during the cold months associated with high cluster mode concentration. Nighttime events have been observed in other Mediterranean studies as well (Carnerero et al., 2018; Kopanakis et al., 2013; Kallioutsi et al., 2012). In a boreal forest, nighttime clustering events that do not usually grow past 5 nm have been attributed to the formation of large highly-oxygenated organic molecules (HOM) mainly from monoterpene oxidation (Lehtipalo et al., 2011; Rose et al., 2018; Bianchi et al., 2019). In the French Landes forest, nocturnal NPF events with clear growth up to 100 nm were attributed to monoterpene oxidation under stratified atmospheric conditions (Kammer et al., 2018). Monoterpenes concentrations reported at the Landes forest reached up to 25 ppb, whereas those measured in the boreal forest were below 2 ppb. Concurrent measurements of biogenic volatile organic compounds (BVOCs) were not available in this study but the average concentration of monoterpenes during March 2015, which is a month with high biogenic activity, was reported to be 0.236 ± 0.294 ppb with a maximum up to 4.5 ppb (Debevec et al., 2018).

Lastly, the EMME region is characterized by a high loading of dust which contributes to around 34% (~10 μg m⁻³) of the annual PM₃₀ levels (Pikridas et al., 2018). In this study, fifty days with high dust loading (Table 4).
S3) were identified based on ground measurements of mineral dust concentrations (Sect. 5 of SI). Among these dates, 37 were NPF event days, 9 were non-events, 2 were undefined and 2 had no data. Figure 10 shows the temporal variation of PM$_{2.5}$, PM$_{10}$, and particle number size distribution measured during three of the dust episodes with ±5 days window before and after the dust episode. NPF took place at high dust loadings, and there is no obvious threshold for the dust loading above which NPF does not occur. In addition, the formation rates (Figure S9) and growth rates (Figure S10) between NPF event days not affected by high dust loading and NPF event days affected by high dust loadings seem to be comparable. $J_1$ was slightly higher on days affected by high dust loading, but this could be related to the lower number of dust cases compared with the non-dust cases. High dust loadings can affect NPF in opposing ways. On the one hand, it can suppress photochemical processes by scavenging reactive gases and condensable vapors (De Reus et al., 2000; Ndour et al., 2009). On the other hand, it can provide particles that can act as a site for heterogeneous photochemistry promoting the formation of gaseous OH radicals, which initiate the conversion of SO$_2$ to H$_2$SO$_4$ (Dupart et al., 2012; Nie et al., 2014). However, a clear association between high dust loading and NPF was not found from the data set presented here.

In this section we analyze two parameters that describe the strength of NPF: particle formation rates ($J_1$) and particle apparent growth rates (GR).

**Particle formation rates:** The particle formation rates for 1.5, 3, and 7 nm particles ($J_{1.5}$, $J_3$, and $J_7$, respectively) were calculated when SMPS measurements were available (until the 2nd of November) and they are presented in Figure 11. $J_{1.5}$ was the highest during the spring; March had the highest median $J_{1.5}$ while April had more events with extreme $J_{1.5}$ values as expressed by the higher mean. In contrast, $J_3$ and $J_7$ did not exhibit a clear seasonality, but their values were generally higher during the spring. The diurnal cycle for the formation rates was more pronounced during the Class I events than during the Class II or bump events, and the peak median hourly value was highest during Class I events (Figure 12). The median peak of $J_{1.5}$ and $J_7$ during the class I events and bump events occurred between 11:00 and 12:00, whereas for $J_3$ the peak occurred between 12:00 and 14:00. For the Class II events the corresponding peaks occurred about 1 hour later. To place the formation rates in global perspective, we compare $J_3$ from this study to other studies (Table 1), because it is the most commonly reported value in literature. The studies in Table 1 were selected on the basis of having one year or more of measurement data. $J_3$ determined in this study were higher than that measured at semi-pristine rural areas (Väisälä, Hyytiälä, and Tomsk), lower than that measured in a megacity (Beijing) but higher than values reported at urban and rural sites affected by urban pollution (Budapest, Helsinki, Vaihulla, and Po Valley).

**Apparent growth rates:** We report size-segregated growth rates between 1.5 and 3 nm (GR$_{1.5-3}$), between 3 and 7 nm (GR$_{3-7}$), and between 7 and 20 nm (GR$_{7-20}$) as recommended by Kulmala et al. (2012) for negatively charged ions, positively charged ions and total particles (charged + neutral) (Figure 13). The growth rates of total particles were higher than that of the charged fraction, which is in agreement with earlier studies showing enhanced growth rates in the neutral channel at diameters below 15 nm (Gonser et al., 2014; Manninen et al., 2009; Rohan Jayaratne et al., 2016). This behavior has been explained by Gonser et al. (2014) who provided a conceptual model of the influence of cluster ion recombination and attachment at different stages of particle nucleation and growth. The seasonal behavior of the growth rates was also distinct. In the sub 3 nm range, the negative ions growth rates had similar median values across the year except during July, which had higher growth rates. The seasonal behavior of the growth rates was also distinct. In the sub 3 nm range, the negative ions growth rates had similar median values across the year except during July, which had higher growth rates.
3.4 The driving atmospheric parameters of the NPF events

To explain the occurrence of NPF at CAO-AMX, we investigated the effect of the following variables: CS, meteorological conditions (temperature, solar radiation, pressure, relative humidity, wind speed and wind direction), trace gas concentrations (NO, SO2, CO and O3), air mass origin, and sulfuric acid. The median of CS at CAO-AMX, for the periods that SMPS measurements were available, was 7.9×10^{-3} s^{-1} (25th-75th percentiles = 5.2×10^{-3} - 13.9×10^{-3}) while the mean was 10.7×10^{-3} s^{-1} ± 8.2×10^{-3} s^{-1} (computed from daily median values). These values lie within the range of coastal (Kalivitis et al., 2019) and urban environments (Salma et al., 2016; Jun et al., 2014). They are higher than the values reported for forests and semi-pristine environments (Dal Maso et al., 2002; Dada et al., 2017), and lower than the values reported for highly polluted cities (Wu et al., 2007). The average diurnal cycle of the size segregated CS for the whole measurement period shows that particles above 50 nm were the main contributors to the CS, even though particles down ~3 nm could also exhibit a high CS (Figure 12). Thus, nucleating aerosols can largely contribute to the available aerosol surface area. The NPF frequency typically decreases with an increasing CS (Pikridas et al., 2012; Salma et al., 2016a; Dada et al., 2017; Dai et al., 2017; Hakala et al., 2019; Hussein et al., 2020). However, NPF has been observed in polluted environments at exceptionally high values of CS, indicating that inefficient cluster scavenging or enhanced cluster growth or a combination of both drives NPF regardless of the high load of pre-existing particles (Yao et al., 2018; Kulmala et al., 2017). In our study, we did not find a clear relation between CS and the monthly NPF occurrence, and NPF did not necessarily occur at low values of CS (Figure 14). To further explore the effect of CS on NPF, we checked whether the NPF event days had lower CS before the onset of nucleation (period from midnight to morning) in comparison to non-event days, but we did not find any apparent association (Figure S13). A possible explanation for why CS is not systematically lower on NPF event days could be similar to that observed at some mountain sites where the sources of NPF precursors and their sinks (i.e., CS) share the same origin, and thus CS is not necessarily a limiting factor (Sellegri et al., 2019).

Next, we inspected the effect of meteorological variables (Figure 15) on the occurrence of NPF. By considering the data from all the months together, NPF events took place over a wide range of meteorological conditions. Higher temperatures seemed to be favorable for intra-monthly NPF occurrence, however the higher growth rates whereas the positive polarity had notable increase in the growth rates in the summer month. The difference in the growth rates at these cluster sizes suggests that the ion induced NPF processes are more important in the positive channel. In the 3-7 nm size range, there was no clear seasonal pattern except that June had the highest growth rates in the negative and positive mode while the month exhibiting the highest growth rates in the total particle mode were February and June. In the 7-20 nm size range, the growth rates exhibited a clear seasonality in all channels with a peak in February and another broad peak during the summer month.

The GR increased with an increasing particle size, which is a typical feature in the sub-20 nm size range because condensational growth is more favorable as the particle size increases and the Kelvin effect decreases (Manninen et al., 2010). The median growth rates in the three size ranges (calculated from the daily means of the three channels) were 3.7, 9.2, and 11.7 nm hr^{-1}, respectively. These GRs are higher than those reported for a rural boreal environment (1.9, 3.8 and 4.3 nm hr^{-1}, respectively) (Yli-Juuti et al., 2011). In comparison to other studies, the ion mode GR reported here is on the higher range of GRs measured at 12 European sites (Manninen et al., 2010 cf. Figure S8). The high growth rates reported here could be associated to the high fraction of bump events. As discussed in Sect. 3.3, these events are characterized by a burst of particles within a short period of time, which would translate to higher growth rates.

Condensation sink: The median of CS at CAO, for the periods that SMPS measurements were available, was 7.9×10^{-3} s^{-1} (25th-75th percentiles = 5.2×10^{-3} - 13.9×10^{-3}) while the mean was 10.7×10^{-3} s^{-1} ± 8.2×10^{-3} s^{-1} (computed from daily median values). These values lie within the range of coastal (Kalivitis et al., 2019).

Condensation sink: The median of CS at CAO, for the periods that SMPS measurements were available, was 7.9×10^{-3} s^{-1} (25th-75th percentiles = 5.2×10^{-3} - 13.9×10^{-3}) while the mean was 10.7×10^{-3} s^{-1} ± 8.2×10^{-3} s^{-1} (computed from daily median values). These values lie within the range of coastal (Kalivitis et al., 2019).
temperatures from June to September did not coincide with higher NPF frequencies (Figure 15a). The effect of temperature on NPF has been studied extensively in chamber experiments, with a general consensus that lower temperatures favor nucleation at the kinetic regime and thus enhance NPF in inorganic systems like the sulfuric acid-ammonia system (Lee et al., 2019). However, in organic systems where highly oxygenated organic compounds (HOM) are the main NPF species, temperature plays a double role. On the one hand, the Gibbs free-energy barrier is reduced at lower temperatures, favoring the condensation of less oxidized vapors that would not condense at higher temperatures. On the other hand, lower temperatures lead to decreased auto-oxidation reaction rates and reduced yields of HOM. Recent studies have shown that the former effect compensates for the latter effect, having an overall increase in nucleation and growth rates at lower temperatures (Stolzenburg et al., 2018; Simon et al., 2020; Ye et al., 2019). While these mechanisms are clear in chamber studies, the situation becomes more complicated in the atmosphere because of the complexity of the atmosphere-biosphere system having simultaneous temperature-dependent processes that can enhance or suppress NPF, making current atmospheric observations inconsistent on the role of temperature on NPF (Kerminen et al., 2018). Solar radiation is regarded as one of the most important factors affecting NPF (Jokinen et al., 2017). Its intensity is relatively high in Cyprus all year round. Intra-monthly, NPF events occurred at higher global radiation during the winter and autumn month, whereas in spring (except April) and summer months, radiation did not seem to be a limiting factor for NPF (Figure 15b). Inter-monthly, the month with the highest solar radiation did not coincide with the highest occurrence of NPF. Regarding ambient relative humidity (RH), NPF events tend to occur at lower RH in both clean and polluted environments (Kerminen et al., 2018). However, high RH values do not necessarily suppress NPF (Salma and Németh, 2019), which agrees with our observations (Figure 15c). In terms of the surface air pressure, intra-monthly NPF was on average observed on days with higher pressures and the inter-monthly NPF occurrence was the lowest during the month with the lowest surface pressure (Figure 15d). With respect to wind speed, high wind speeds did not seem to prevent NPF but event days occurred mostly under low wind speeds (Figure S14). In terms of wind direction, NPF occurred mainly when the wind was blowing from the west to east sector but with a frequency of occurrence which is higher in north to easterly winds (Figure S14, S15, & S16). The north-to-easterly direction is the direction where the main agglomerations and livestock farming lands are situated. These local sources could be enhancing the occurrence of NPF but no direct relation was found between the north-to-easterly wind direction and specific event types (Figure S16). However, NPF class I events did not occur when the wind was originating from the southeast to south west sector.

The EMME region is among the regions with the highest background of trace gases and aerosols concentrations in the Northern Hemisphere (Lelieveld et al., 2002). Here, we investigate the relation between trace gases and the occurrence of NPF (Figure 15a). The intra-monthly SO$_2$ concentration was, on average, higher during NPF event days in comparison to non-event days during most of the months. The inter-monthly SO$_2$ concentrations during the highest NPF occurrence (Mar-May) were similar to the months with the lowest NPF occurrence (Jun-Aug). This indicates that SO$_2$ and thus sulfuric acid (as will be shown subsequently) cannot alone explain the seasonal pattern of NPF. The ozone (O$_3$) concentration is particularly high in Cyprus and is mainly influenced by regional and transported ozone, while local precursor emissions play a minor role in ozone formation (Kleinthous et al., 2014). Intra-monthly, the O$_3$ concentration was sometimes lower, similar or higher during NPF event days compared with non-event days, with no clear seasonality. One notable remark is that in April, NPF events took place at much higher O$_3$ concentrations than what was observed on non-event days. This could imply that higher oxidative capacity was driving NPF during April. April also had the most notable differences in global radiation and RH between NPF event and non-event days. Analogous to SO$_2$, O$_3$ cannot explain the seasonal pattern of NPF. CO levels are generally high over the Mediterranean (in comparison to the Pacific), with emission sources being typically from western and eastern Europe, having
lower contribution from the regions surrounding the Mediterranean (Lelieveld et al., 2002). The CO - NPF relationship at CAO_AMX did not have a distinct character, however CO concentrations were slightly lower during the summer months. Regarding NOx, NPF event days had on average higher NOx concentrations within the boundaries of each month, except in April. More notably, NOx had lower concentrations during the months with lower NPF frequencies, which might indicate the role of associated anthropogenic organic vapors in triggering NPF at our site.

We examined the effect of air mass origin arriving at CAO_AMX at 8:00 a.m. during event and non-event days from seven source regions: local, N. Africa, marine, Europe, Asia, NW. Asia, and SW. Asia (Figure 17). The last two source regions (NW. Asia, and SW. Asia) represent the geographic location with respect to Cyprus location. An obvious feature that pops out is that the months with the highest NPF frequency had air masses originating from south-west Asia (the Middle East), whereas during the month with the lowest NPF frequency air masses did not originate from that direction. This pattern might suggest that chemical compounds important for nucleation and subsequent growth are transported to CAO_AMX from the Middle East. Between the end of spring and late September, which are the months with the lowest NPF frequency, the circulation over the eastern Mediterranean is characterized by persistent northerly winds called the Etesians (Tyrili and Lelieveld, 2013). The NPF events during this period, as shown in Sect. 3.2, were weak or generally did not lead to particle growth into large sizes in comparison to the rest of the year. The Etesian circulation flow is caused by a sharp surface pressure difference between the westerly Azorean high-pressure regime and the Asian monsoon low-pressure regime. While the Etesians block the north-ward transport of desert dust, they trigger high sea levels, prevent rain over the region, and enhance marine inversions (Ulbrich et al., 2012). They favor the transport of air pollutants from Central/Eastern Europe and west Turkey and, together with enhanced photochemical conditions and low precipitation, contribute to high O3 (Solomou et al., 2018) and particulate matter (PM) levels (Pikridas et al., 2018). The increase in PM levels during these months could be a limiting factor for NPF. Since we did not have particle size distribution measurements above 700 nm, therefore bigger particles could be additionally contributing to the CS. However, from a mass concentration point of view, PM2.5 and PM10 did not show a pattern that would support this hypothesis (Figure S17 & S18). Additionally, while the south-west Asia sector might be important for NPF, it did not exhibit a clear pattern during the month with the highest NPF frequency. In fact, in April most of the air masses originated from north-west Asia. This sector appears to be also important for NPF during Jun, July and November, whereas the other sectors did not exhibit any notable pattern. Air masses did not originate from the Asia sector because they were obscured by the terrain height. Pure marine air masses were not as frequent as other air masses. However, owing to the location of Cyprus, air masses from other continental source origins are expected to have been influenced by marine conditions as they travel to our measurement site. Thus we cannot exclude the potential marine effect on the occurrence of NPF.

Last, we investigated the role of sulfuric acid. While sulfuric acid (H2SO4) is considered as one of the main nucleating species in the atmosphere, it is well known that H2SO4 binary nucleation with water requires high H2SO4 vapor concentrations that are not relevant within the lower parts of the troposphere (Wyslouzil et al., 1991). Additional species are required to stabilize H2SO4 clusters, such as ammonia, amines or ions, while some other compounds can nucleate on their own in atmospherically relevant conditions, including iodine oxides and highly oxygenated organic compounds (HOM) from biogenic precursors (Lee et al., 2019 and references therein). In this study, the hourly H2SO4 proxy concentrations ranged between 3×10^5 and 1×10^6 cm^-3 which are typical values for H2SO4 in the troposphere. The relationship between particle formation rates (J(t)) and H2SO4 proxy concentration varied across the month of the year (Figure 18). Lower concentrations of H2SO4...
were required during winter and spring to achieve the same formation rates as in the other seasons. A possible explanation to this behavior is that in the first case, stabilizing compounds are abundant in the atmosphere and thus H$_2$SO$_4$ is required for the formation of particles. A similar hypothesis was tested by Pikridas et al. (2012) by using the accumulation mode particle acidity as an indirect measure of the availability or lack of ammonia or any other basic species in the gas phase. The authors concluded that excess base is not available during the summer to participate in the nucleation process. In our case, however, the formation rate versus H$_2$SO$_4$ relationship is closer to those derived for the H$_2$SO$_4$-DMA-NH$_3$-H$_2$O system than those for the H$_2$SO$_4$-NH$_3$-H$_2$O system. In fact, the ternary nucleation of H$_2$SO$_4$-NH$_3$-H$_2$O is unlikely to be important at ground level because of too low concentrations or too high temperatures (Kurten et al., 2018). This suggests that, in our case, the missing stabilizing base is probably not ammonia, although the role of ammonia cannot be ruled out. The distinct air mass origin during the summer could explain the decrease in the concentrations of the stabilizing base. Otherwise, the high temperature during the summer could be the factor that disfavors the occurrence of NPF. Most certainly, NPF at CAO-AMX seem to be influenced by several factors and chemical constituents. This has been also indicated by Debevec et al. (2018) whom observed four types on nucleation events, within one month of measurements, having: 1) predominant anthropogenic influence, 2) predominant biogenic influence, 3) mixed anthropogenic - biogenic influence, and 4) a marine influence with low concentrations of anthropogenic and biogenic tracers. Therefore, to reveal the main mechanisms of NPF, long-term measurements of nucleating clusters and organic precursors using state-of-the-art online mass spectrometry techniques are essential.

3.5 Regression and classification analysis

To further understand the occurrence of NPF events, we present in the section 12 of the SI material, two types of analysis: the first is a linear regression analysis of formation rate of 1.5 nm particles ($J_{1.5}$) and the second is a decision tree classification model to indicate whether each day is an NPF event day or a non-event day. Both analyses have shown that NO$_2$, H$_2$SO$_4$ and wind direction (mainly from N-to-E direction) are the most important parameters that are associated with NPF occurrence (Figures S20-S23). While the role of H$_2$SO$_4$ in NPF is well known in literature, the role of NO$_2$ is not that clear. NO$_2$ has been shown to play contrasting roles in NPF depending on the associated pool of gas molecules. On the one hand, when oxidized to nitric acid, it can enhance NPF in the presence of ammonia vapors (Wang et al., 2020). On the other hand, it can suppress NPF by reducing autoxidation and low-volatility HOM dimer formation (Wildt et al., 2014; Zhao et al., 2018). Nevertheless, Yan et al. (2020) have shown that this effect is weak when NH$_3$ and H$_2$SO$_4$ are additionally present and that NO$_2$ is more effective than NO$_2$ in changing the HOM composition and volatility. Xie et al. (2015) have revealed that NO$_2$ can play an important role, not only in surface catalytic reactions of SO$_2$, but also in dust-induced photochemical heterogeneous reactions of NO$_2$, which produces additional sources of OH radicals and promote new particle formation and growth. However, while NPF seems to occur more frequently at higher NO$_2$ concentrations in our study, we cannot conclude if it plays a role in NPF or if it is a proxy of some other pollutant, especially that NO$_2$ concentrations were mostly lower than 4 ppb. What is evident however is that H$_2$SO$_4$ does not nucleate on its own at the concentrations reported in this study, thus an unknown stabilizer and possibly other compounds participating in NPF are missing in this analysis (as explained in the regression analysis). We hypothesize that these unknown compounds (e.g. NH$_3$ / amine /HOM) are associated with the North to east wind directions and higher NO$_2$ concentrations.

4 Conclusion

Recent studies have pointed out that NPF is important in the EMME region (Brilke et al., 2020; Debevec et al., 2018; Hakala et al., 2019; Hussein et al., 2020; Kalivitis et al., 2019; Kalkavouras et al., 2019). Brilke et al. (2020) studied NPF in a coastal site in Cyprus with strong local pollution during 2017, while Debevec et al.
(2018) characterized NPF at the same site of this study during 2015. While both studies were limited to one month of observations, we disclosed here the first long-term (one year) characterization of NPF at a background site in Cyprus. We presented the general and seasonal characteristics of PSD and NPF then we explored the factors that affect NPF.

Our analysis of NPF intra-monthly variability showed that on average, NPF events occurred at higher temperatures, lower RH and higher global radiation, except during the months of August, September and December. To the contrary, lower pressure conditions, higher wind speeds and local south to west wind directions seemed to be more favorable for non-events. The frequency of NPF was higher than that reported at a similar Eastern Mediterranean island site using a slightly limited measurement setup than the one applied here. This demonstrates the importance of comprehensive measurements using instruments that can measure down to cluster sizes. NPF occurred all year round, with higher frequencies during the spring and autumn and a minimum frequency during the summer. The particles did not grow significantly after nucleation during the months with the lowest NPF frequencies. These months were also characterized by lower NOx concentrations, an indication of lower anthropogenic influence, and distinct air mass origin profiles from the rest of the year.

Condensation sink, calculated based on a PSD up to 700 nm, had no clear relationship with NPF, but it was slightly higher during some summer month. Additionally, sulfuric acid was not the limiting factor for NPF seasonality as its estimated concentration was mostly high during the summer, up to 1e7 molecules cm⁻³. The relationship between particle formation rates and sulfuric acid proxy exhibited different slopes between the months with the highest and lowest NPF frequency, suggesting that nucleation might have proceeded with varying temperatures or at different concentrations of stabilizing compounds and other aerosol precursors not measured in this study.

The analysis presented in this study is a step forward towards understanding the mechanisms of NPF mechanism in the EMME region. Future studies require long-term measurements of vapors that participate in NPF and subsequent growth. These could include, for example, ultra-low volatility organic compounds (ULVOC), extremely low-volatility organic compounds (ELVOC), low-volatility organic compounds (LVOC), ammonia, amines and iodic species. Further, to understand the ubiquity of the effect of large particles which could inhibit NPF during certain episodes but enhance NPF during episodes with high mineral dust loadings, extended PSD measurements up to coarse particles, preferably coupled with chemical speciation, are important. On a larger scale, long-term measurements of CCN particles are necessary to quantify the contribution of NPF to the CCN budget. These measurements would preferably take place not only in Cyprus but also in different location in the Middle East and North Africa.
Data availability

The data used in this study is available at https://doi.org/10.5281/zenodo.4701303.

Author Contributions

The study was conceived by MK and JS. RB, TJ, TL, KN and MP prepared and installed the instruments. RB, MP, KN, AM, EB, AC, and FU performed the regular maintenance for the instruments. RB performed the data analysis and wrote the manuscript. LD provided support in data analysis. LA provided support in instrument troubleshooting and nCNC inversion. MP and EB performed the sector analysis. SB performed the kappa measurements and provided support on hygroscopicity calculations. MP and JS provided the SMPS data. CS provided the meteorological and trace gas data. RB, MP, TJ, LD, SB, KL, JK, GB, TP, VMK, and MK participated in the scientific discussion and reviewed the manuscript.

Competing interests

The authors declare no conflict of interest.

Financial support

This publication has been produced within the framework of the EMME-CARE project which received funding from the European Union’s Horizon 2020 Research and Innovation Programme, under Grant Agreement No. 856612 and from the Cyprus Government. This work has received additional funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (ERC, Project No.742206 “ATM-GTP”). The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein. Additional support was received from the Academy of Finland (Grant Agreements No. 307331, 337549, 302958, and 316114), the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (Project: INTEGRATED/0916/0016). JK acknowledges support from Academy of Finland (project 1325656) and University of Helsinki 3 year grant 75284132. TJ acknowledges support from Academy of Finland (project 334514).

Acknowledgments

The authors thank Hanna Manninen, Kaspar Dällenbach, Jenni Kontkanen, Runlong Cai and Dominik Stolzenburg for fruitful scientific discussions. Frans Korhonen, Pekka Rantala, Pasi Aalto, Erki Siivola, Sander Mirme, Joonas Vanhanen, and Aki Halonen, are kindly acknowledged for their continuous and indispensable support.
Table 1. Comparison of formation rates from long-term measurements at various sites including this study

<table>
<thead>
<tr>
<th>Site</th>
<th>Type</th>
<th>Period</th>
<th>$J_f$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAO-AMX, Cyprus</td>
<td>Rural</td>
<td>2018</td>
<td>6.42±8.47</td>
<td>4.19</td>
</tr>
<tr>
<td>Vavihill, Sweden</td>
<td>Rural</td>
<td>2001-2004</td>
<td>4.3</td>
<td>1.89</td>
</tr>
<tr>
<td>Hyytiälä, Finland</td>
<td>Rural</td>
<td>1996-2004</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Värriö, Finland</td>
<td>Rural</td>
<td>1996-2004</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Tomsk, Russia</td>
<td>Rural</td>
<td>2005-2006</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Beijing, China</td>
<td>Urban clean Urban polluted</td>
<td>2004-2005</td>
<td>22.3±15.1</td>
<td>18.7</td>
</tr>
<tr>
<td></td>
<td>Urban</td>
<td>2001-2003</td>
<td>17±20</td>
<td>9</td>
</tr>
<tr>
<td>Po Valley, Italy</td>
<td>Rural w. urban influence</td>
<td>2002-2005</td>
<td>5.89</td>
<td>3.31</td>
</tr>
<tr>
<td>Budapest, Hungary</td>
<td>Urban</td>
<td>2008-2009</td>
<td>4.2±2.5 a</td>
<td>4.2 a</td>
</tr>
<tr>
<td>Helsinki, Finland</td>
<td>Urban</td>
<td>1997-2006</td>
<td>2 b</td>
<td>1.09 b</td>
</tr>
<tr>
<td></td>
<td>Arctic</td>
<td>2016-2018</td>
<td></td>
<td>0.02-1.62</td>
</tr>
<tr>
<td>Finokalia, Greece</td>
<td>Remote coastal</td>
<td>2008-2009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) data represent 6-25 nm range

\(^b\) calculated from monthly values
Figure 1. Maps of the Mediterranean region and Cyprus. (a) Location of Cyprus in the Mediterranean region. (b) Location of the measurement site (CAO-AMX) in Cyprus. (c) Location of the measurement site (CAO-AMX) pointed by the blue location marker with respect to the villages of Agia Marina and Xyliatos. The geographic border of the villages is marked by the yellow enclosure. The maps were retrieved from Google (©2020 Google, TerraMetrics).
Figure 2. Monthly variation (at radiation >50 W. m$^{-2}$) of particle number concentration of (a) cluster mode, (b) nucleation mode, (c) Aitken mode, and (d) accumulation mode presented by box plots. The central red marks indicate the median, the blue small boxes indicate the mean, the bottom and top edges of the big box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol. Data presented have daily time resolution. Month designated with '*' symbol have less than 20 days of data. Note that SMPS measurements were not available on November and December.
Figure 3. The diurnal cycle (at radiation > 50 W m⁻²) of particle number concentration of Cluster mode (a), Nucleation mode (b), Aitken mode (c), and Accumulation mode (d). The shaded areas with black dashed boundaries represent the 25th and 75th percentile limits while the solid line represents the median and the squares indicate the mean. Notice the difference in the y-scale between the top and bottom plots.

Figure 4. Classification of NPF events presented (a) annually and (b) monthly.

Deleted:
Figure 4. Calendar of daily event day classification between January 27, 2018 – January 26, 2019 and the monthly fraction of NPF event classes.
Figure 5. Monthly percentage of the classes of event days.

Figure 6. Percentage histograms showing the frequency distribution of (a) NPF events growing to a certain diameter, (b) NPF event duration, and (c) the event start and end times from sunrise.
Figure 7. Examples of days with multiple nucleation events: (a) March 18, 2018 (b) March 22, 2018 (c) April 8, 2019, (d) April 30, 2018, (e) September 9, 2018 and (f) September 30, 2018.
Figure 8. Example of days showing a decreasing mode diameter: (a) March 27, 2018, (b) April 1, 2018, (c) April 18, 2018, (d) May 2, 2018, (e) September 17, 2018 and (f) October 27, 2018 in local time. The mode diameter is plotted as black circular markers.

Figure 9. Examples of days with nighttime clustering and growth marked with white rectangles: (a) February 7, 2018, (b) February 22, 2018, (c) March 21, 2018, (d) October 20, 2018, (e) December 04, 2018 and (f) December 18, 2018. The black line is the solar radiation (W.m$^{-2}$) which can be read from the right axis. Note the difference in the color scale used in this figure in comparison to figures 6 and 7.
Figure 10. Temporal variations of aerosols during dust episodes with 5 days before and 5 days after the dust episode. (a) Time series of particle size distribution, PM$_{10}$, and PM$_{2.5}$ between Feb 1, 2018 and Feb 15, 2018 (dust episode: Feb 6 to Feb 10). (b) Time series of particle size distribution, PM$_{2.5}$ (coarse PM), and PM$_{2.5}$ between Mar 15, 2018 and Apr 2, 2018 (dust episode: Mar 20 to Mar 28). (c) Time series of particle size distribution, PM$_{10}$, and PM$_{2.5}$ between Apr 26, 2018 and May 15, 2018 (dust episode: Apr 26 to Apr 27 and May 1 to May 7).
Figure 11. Monthly variation of particle formation rates during NPF events: (a) $J_{1.5}$, (b) $J_3$ and (c) $J_7$. The central marks indicate the median, the blue small boxes indicate the mean, the bottom and top edges of the big box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the ‘+’ symbol. The numbers above the box plot represent the number of data points within each boxplot. Data presented have daily time resolution. Daily J values were calculated by taking the mean of hourly values within event duration times.
Figure 12. Diurnal variation of J1.5 (top), J3 (middle) and J7 (bottom) during class I, class II, bump events and non-events. Shaded areas represent the 25th and 75th percentile bounds while the solid line represents the median.
Figure 13. Monthly variation of growth rates during NPF events in three size ranges: (a) <3nm, (b) 3-7nm, and (c) 7-20nm. The central marks indicate the median, the bottom and top edges of the big box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol. The numbers above the box plot represent the number of data points within each boxplot. Black boxes represent the total particles (neutral+charges), blue boxes represent negative ions and red boxes represent positive ions.
Figure 14. Monthly variation of condensation sink (s$^{-1}$) during event (blue) and non-event (green) days using data corresponding to global radiation is greater than 50 W.m$^{-2}$. The bottom and top edges of the box plots indicate the 25th and 75th percentiles, respectively. The central mark indicates the median. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol. Data presented have hourly time resolution. The shaded grey bars represent the monthly NPF percent occurrence.
Figure 15. Monthly variation of metrological parameters during event (blue) and non-event (green) days: (a) temperature, (b) global radiation, (c) relative humidity, (d) pressure, (e) wind speed, and (f) wind direction using data corresponding to global radiation greater than 50 W.m$^{-2}$. The bottom and top edges of the box plots indicate the 25th and 75th percentiles, respectively. The central mark indicates the median. The whiskers extend to the extreme data points not considered outliers. Data presented have hourly time resolution. The shaded grey bars represent the monthly NPF percent occurrence.

Figure 16. Monthly variation of trace gases during event (blue) and non-event (green) days: (a) sulfur dioxide, (b) ozone, (c) carbon monoxide, and (d) nitrogen oxide using data corresponding to global radiation greater than 50 W.m$^{-2}$. The shaded grey bars represent the monthly NPF percent occurrence. The bottom and top edges of the box plots indicate the 25th and 75th percentiles, respectively. The central mark indicates the median. The whiskers extend to the most extreme data points not considered outliers. Data presented have hourly time resolution.
Figure 17. a) The source regions of air masses reaching CAO-AMX used for the air mass sector analysis. The map was plotted using The Climate Data Toolbox for MATLAB (Greene et al., 2019). The IHO World Sea Areas v3 were used to retrieve the boundaries of the Mediterranean Sea (Flanders Marine Institute, 2019). Note that marine areas other than the Mediterranean Sea were considered part of the continental sectors and that the NW. Asia and SW. Asia sectors are with respect to Cyprus location. b) Monthly variation of air mass origin arriving at CAO-AMX at 8:00 a.m. during event (E) and non-event days (NE). There are no air masses originating from Asia sector because those are obscured by terrain height.

Deleted: 13
Figure 18. $J_{1.5}$ versus sulfuric acid proxy concentrations color-coded by the month of the year. Data presented have hourly time resolution.
References


Kurten, T., Loukonen, V., Vehkamäki, H., and Kulmala, M.: Amines are likely to enhance neutral and ion-induced sulfuric acid-water nucleation in the atmosphere more effectively than ammonia, Atmospheric Chemistry and Physics, 8, 4095-4103, https://doi.org/10.5194/acp-8-4095-2008, 2008.


