Hirshorn et al., Responses to Editor

The authors of this paper would like to thank the editor for reviewing the reviewer responses and providing suggestions on how to prepare the paper for publication. We have carefully considered the feedback. Based on this feedback, we have made appropriate changes to address remaining concerns. Please note that references to line numbers below in the author responses correspond to the new line numbers in the updated manuscript.

The authors have color coded the responses to the reviewer as follows:

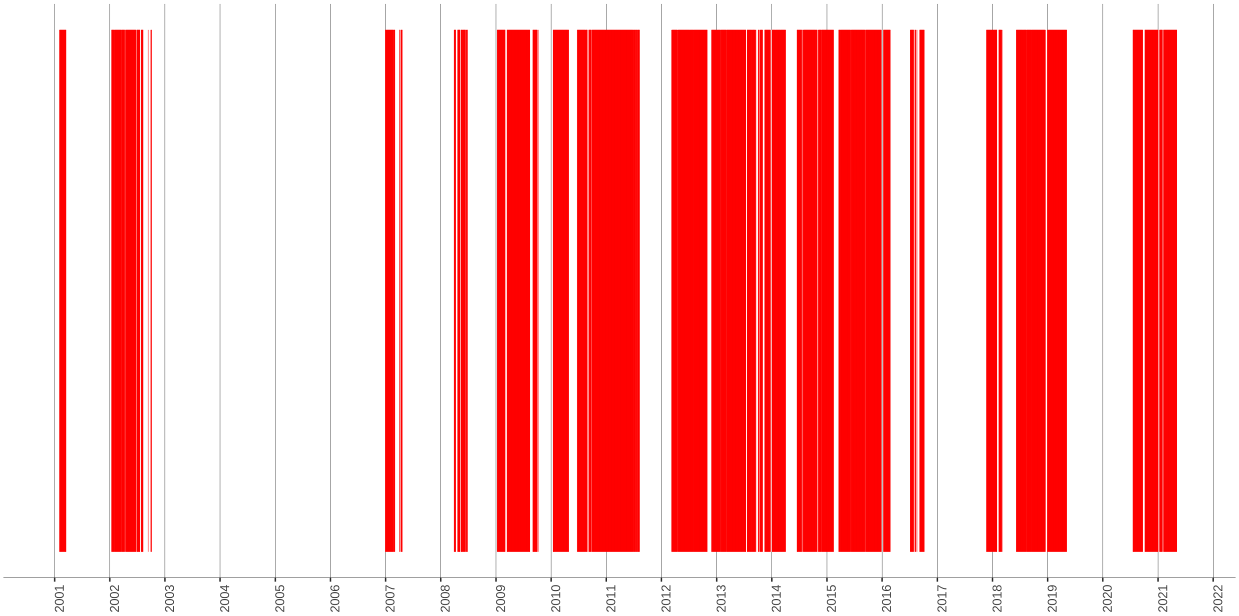
Blue: A response to comments provided by the editor/reviewer.
Black: Text that is in the originally submitted manuscript.
Red: Changes that were made to the manuscript and are reflected in the updated manuscript.

1. About comment#2 and authors response.
a) We can agree that statistical methods are better than visualization method, however if the visualization method has been the standard, better comparison is needed and in section 5 you need to extend your analysis/discussion. Actually, the figure 6 presents 2006-2021 comparison with just n=835 days. Have the authors used ~3years data in a period of 15 years, why then the figure have 2006-2021 title? In addition, the advantage against other automatic methods presented in the answers is not really clear for the reviewer, but at least some discussion need to be included in the manuscript.

Addressing figure 6, the authors would like to acknowledge that the data considered in this study spans 2006-2021 which is why the label is included in the title. There are periods of time when the SMPS was offline due to maintenance or an instrumental issue and thus data was not collected to ensure quality. Furthermore, some days are removed from consideration if they do not meet the data quality requirements described in section 2.2 of the paper:

Lines 139-144: The first step of the automatic classification method is to ensure the availability of SMPS level 1 data. Although NPF events can span multiple days, we consider daily data (0:00 – 23:59 MST) as well as the first 12 hours (0:00 – 12:00 MST) of the next day to ensure the consideration of an NPF event doesn't prematurely end if growth continues overnight. 5-minute SMPS data is only considered if the first 24-hour period meets the following conditions: there are at least 16 hours of data present, and the period between 10:00 – 23:00 MST (the times in which NPF is most common at SPL) has less than 1 hour of data missing.

In addition, the authors would like to provide a plot that details the availability of SMPS data over the course of the study:



Regarding the additional comparison to other methods, the authors agree that an in-depth comparison will add to the paper and needs to be a subject of future work. To get access to automatic methods (Joutsensaari et al., 2018; Su et al., 2022) based on machine learning the authors would need to train methods using data in this study. As a result, we would be unable to reuse the data. An additional comparison paragraph was added to the paper:

Lines 446-454: While a comparison with the automatic methods that use deep learning based convolution neural networks (CNN) (Joutsensaari et al., 2018; Su et al., 2022) would provide an important comparison, training the CNN would require the removal of the data used in training from consideration. For example, Su et al., 2022 requires 358 annotated days to train and only classifies class 1 (banana shaped) events while our method can also identify class II days. Joutsensaari et al., 2018 presents another option of automatic classification using deep learning but recommends 150 days per class to properly train the method for each site. The big advantage of our method compared to other automatic methods is that aspects of the statistical method can be altered to fit individual sites without having to train the method. Assuming there are enough data available, future studies focusing on using automatic methodology should attempt to use both the statistical method detailed here, and CNN based automatic methods.

b) The reviewer acknowledge the answer but can not agree. In order to maintain consistency with a simplified/wrong equation, the authors can use that formula to compare with Hallar et al. (2011) but not to provide new data. The equation used by the authors is used in Hallar et al. (2011) and Kulamala et al. (2004), however this formula does not account for losses. The authors stated that "because of the clean conditions at SPL" they keep using the Hallar et al. (2011) formula, however, the equation for formation rate (Kulmala et al., 2012) does not really depend on clean or not clear conditions of the site, depends of the losses by coagulation, condensation and instrumental losses (this term we can omit). Check Kulamala et al. (2012) for equations. For clean environments the GR

factor of the formation rate could even be larger than the factor ΔN/ΔDp. I will not accept a manuscript presenting a new methodology that uses an old/wrong/simplified formula. Same for α, i will only accept the value of unity just in case the authors demonstrate it has a minor impact on the CS values (not just because to be consistent with Hallar et al. 2011).

J8 Values:

Before responding to this comment the authors would like to thank the reviewer for pointing out the necessity of including loss terms in the equation and thank the editor for allowing us the time to properly calculate these terms for our paper.

In Kulmala et al., 2012, the following equation is used to address the particle formation rate (J_{d_p}):

$$J_{d_p} = \frac{dN_{d_p}}{dt} + CoagS_{d_p} * N_{d_p} + \frac{GR}{\Delta d_p} * N_{d_p} + S_{losses} \quad (1)$$

$\frac{dN_{d_p}}{dt}$ denotes the time evolution of the particle number size concentration, N_{d_p}. The rest of the terms highlight the relative losses for aerosol formation. $CoagS_{d_p}$ is the coagulation sink in the size range of d_p, and $\Delta d_p + d_p$, where d_p is defined as the aerosol diameter and Δd_p is the range in particle diameters. GR is the particle growth rate (in nm/s and S_{losses} are additional losses. Since S_{losses} encompasses losses, such as instrument losses, that are negligible in observational work, this term is the one term of the Kulmala et al., 2012 equation that is not included in our calculation (Casquero-Vera et al., 2020).

We use this equation to calculate a particle formation rate at an aerosol diameter of 8 nm (J_8) using an aerosol diameter size range (Δd_p) of 8 – 25 nm. To calculate the additional loss terms, Lehtinen et al., 2007 created a relationship relating the condensation sink (CS) and the $CoagS_{d_p}$ using a power-law dependence that is used in Kulmala et al., 2012:

$$CoagS_{d_p} = CS * \left(\frac{d_p}{0.71}\right)^m \quad (2)$$

Where m is a constant equal to -1.6. Moreover, to appropriately address the $CoagS_{\Delta d_p}$ term in equation (1), a diameter size range (Δd_p) is introduced to equation (3):

$$CoagS_{\Delta d_p} = CS * \left(\frac{\Delta d_p}{0.71}\right)^m \quad (3)$$

We consider the whole particle size distribution (8 nm – 333.8 nm) for the CS and coagulation sink to more accurately account for potential particle losses. By using the new equation, which we have high confidence in this method, we calculate new values of J8 for the paper and address these changes in the methodology:

Additions lines 214-229: The J_8 value for an event is defined by the formation rate equation (Kulmala et al., 2004; Kulmala et al., 2012):

$$J_8 = \frac{\Delta N_{8,Dmax}}{\Delta t} + CoagS_{d_p} * N_{d_p} + \frac{GR}{\Delta d_p} * N_{d_p} \quad (3)$$

Where $\Delta N_{8,Dmax}$ is the change in the number concentration of particles across the considered size distribution from about 8 nm to 25 nm ~~the maximum diameter (about 340 nm)~~ during Δt which is the time difference from the defined start of an event to the defined end of an event. When calculating the initial and final number concentrations, we utilize the average number concentration observed between 4 hours and 1 hour prior to NPF initiation as the initial number concentration. The final number concentration is the average number concentration from all 5-min scans taken during an event. Doing so allows for the comparison of the initial conditions of an NPF event, and aerosol formation across the entirety of a given event. The additional loss terms in the equation represent loss to the coagulation sink, and loss due to growth out of the size range (Kulmala et al., 2012). The entire size distribution measured by the SMPS is used when calculating the coagulation sink loss term (Casquero-Vera et al., 2020). ~~We use the above formation rate equation because conditions at SPL are conducive to clean, homogenous air masses allowing for the use of the simplified version of the equation (Kulmala et al., 2004; Hallar et al., 2011).~~

Lines 431 – 436: Average seasonal J_8 values range from 1.76 $\#/cm^{-3}\ s^{-1}$ to 11.07 $\#/cm^{-3}\ s^{-1}$, which are higher than the average seasonal values observed at SPL in 2011 ranging from 0.37 $\#/cm^{-3}\ s^{-1}$ to 1.19 $\#/cm^{-3}\ s^{-1}$ (Hallar et al., 2011). Because this study uses the methodology of Kulmala et al., 2012 and Haller at al., 2011 uses methodology from Kulmala et al., 2004, differences between the two studies are expected since loss terms are not considered in the simplified equation used in Hallar et al., 2011 (Kulmala et al., 2004; Hallar et al., 2011; Kulmala et al., 2012). ~~Because the determination of start and end times differs between visual and automatic classification methods, these lower J_8 values may have to do with the longer time considered in an NPF event.~~

Edits to Table 2:

Average Formation Rate (J_8) ($cm^{-3}\ s^{-1}$)	~~0.23 ± 0.22~~ 3.51 ± 4.35	~~0.33 ± 0.51~~ 11.07 ± 22.35	~~0.12 ± 0.15~~ 1.86 ± 3.14	~~0.17 ± 0.21~~ 1.76 ± 2.41

CS: Mass accommodation coefficient:

The mass accommodation coefficient (α) that is used in this study, is highlighted in the Fuchs–Sutugin correction coefficient (β_m) within the condensation sink (CS) equation in Tuovinen et al., 2021. β_m is depicted as:

$$\beta_m = \frac{1+Kn}{1+\left(\frac{4}{3\alpha}+0.337\right)Kn+\frac{4}{3\alpha}Kn^2} \quad (4)$$

In Kulmala et al., 2012, the following equation for the Fuchs-Sutugin correction coefficient (β_i) was used:

$$\beta_i = \frac{1+Kn}{1+1.677Kn+1.333Kn^2} \quad (5)$$

It is important to note that equation (6) used in Kulmala et al. 2012, is the same as equation (5) that is in Tuovinen et al., 2021, but with the assumption of α = 1. Thus, we continue to assume that the value for mass accommodation coefficient (α) is unity, in this study.

Additional changes: Both Figures 2 and 3 are edited to show accurate J8 and CS values.

c) The authors provide on comment#2 a figure that presents a Gaussian of dN/dlogD vs Dp. However, the comment to L127-136 (page 11 of authors response) indicate that the Gaussians are time dependent? Please clarify if for the GR the authors are using the "Maximum-concentration method" or the "Log-normal distribution function method" (Kulmala et al. 2012). How you choose the start and end time/diameters of the Gaussians

The authors want to clarify that the figure in comment two is a conceptual figure used to illustrate how the calculations of a single Gaussian at a given size bin occur. The Gaussian calculations find the given time at which the maximum Gaussian occurs at a single size bin. The authors detail this calculation of individual Gaussians in the following portion of the manuscript on lines 153 - 164:

The normal distributions were fit by solving for the non-linear least-squares estimates using the R programming language (Equation 1) which considers the particle size distribution at each diameter to return the time that corresponds to the maximum concentration at that given diameter (Bates and Watts, 1988). In the equation, "k" is the maximum aerosol number concentration, "t" is the time index where the normalized maximum at D_p occurs, "μ" is the mean aerosol concentration, and "σ" is the corresponding standard deviation. This equation is used for the calculation of individual maximum Gaussians at each size bin:

$$f(t \mid k, \mu, \sigma) = ke^{-\frac{(t-\mu)^2}{2\sigma^2}}, \ \ k = max\left(\frac{dN}{dlogD_p}\right) \quad (1)$$

The derived time index represents the time at which the maximum of the peak fitted particle size distributions occurs for each value of D_p. For data where at least 5 different Gaussian maximum points are calculated, a linear regression is fit to these maxima allowing for further analysis of growth over the course of an event (Lehtinen and Kulmala, 2003).

Once the Gaussians are calculated, the growth rate can be determined based on where the Gaussians are positioned related to time. This method is most similar to the log-normal function distribution method of calculating growth rate but fits the growth rate using the position of the maximum gaussians.

Lines 181 – 183: Because the slope of the linear regression fit of the maximum Gaussians represents particle growth over time during NPF events, this value is used when determining the growth rate. This method is most similar to the log-normal function fitting method of calculating growth rate but finds the growth rate by fitting a linear regression to the maximum Gaussians.

Page 8 on answers' document: "Thus, the methodology within this paper carefully considers similar timeframes within the diel pattern with and without NPF, to look at the relative change induced by NPF". You consider time frames for event and non-event days, but background conditions are the same? The occurrence of NPF at mountain sites is triggered by an increase condensable vapors that usually is accompanied of an increase on particle concentrations. Thus, particle concentrations (and CCN) usually is larger during NPF events, not because this particles come

Thank you for providing this clarification, the authors better understand what the editor and reviewer are asking. When crafting the answer on page 8, the authors did consider two aspects detailed within the work of Sellegri et al., 2019. First, at high altitudes over 1,000 meters, transport of condensable vapors can be accompanied by particles. However, this same study lists SPL as an exception where NPF is associated with low-surface area of pre-existing particles. To analyze this relationship, the condensation sink in this study is calculated for times before NPF initiation during events, and similar representative times during non-events.

To address that the transport of pre-existing particles could be an error, the following was added to the manuscript:

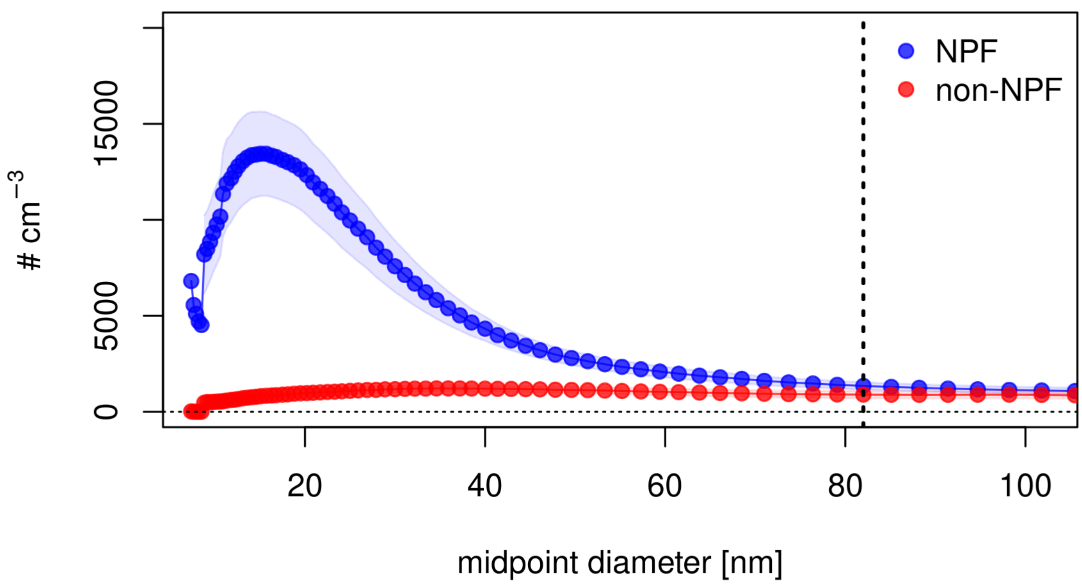
Lines 285 – 289: At other high-altitude mountaintop sites around the globe, this approach could have sources of error since NPF can be associated with the transport of both condensable vapors and pre-existing aerosol that could become CCN (Sellegri et al., 2019). However, SPL seems to be an exception to this rule since previous observations of NPF show association with lower existing particle surface areas which allows for a more direct comparison of events and non-events (Hallar et al., 2011; Sellegri et al., 2019).

Lines 380 – 383: Because the CS is calculated before NPF initiation, these trends further suggest that aerosol transport to the site is not affecting the background particle concentrations during events. More work to analyze the relationship between CS and particle transport is required since the role the CS has on NPF is highly dependent on the conditions of a given site ~~which is why it is important to report CS values~~.

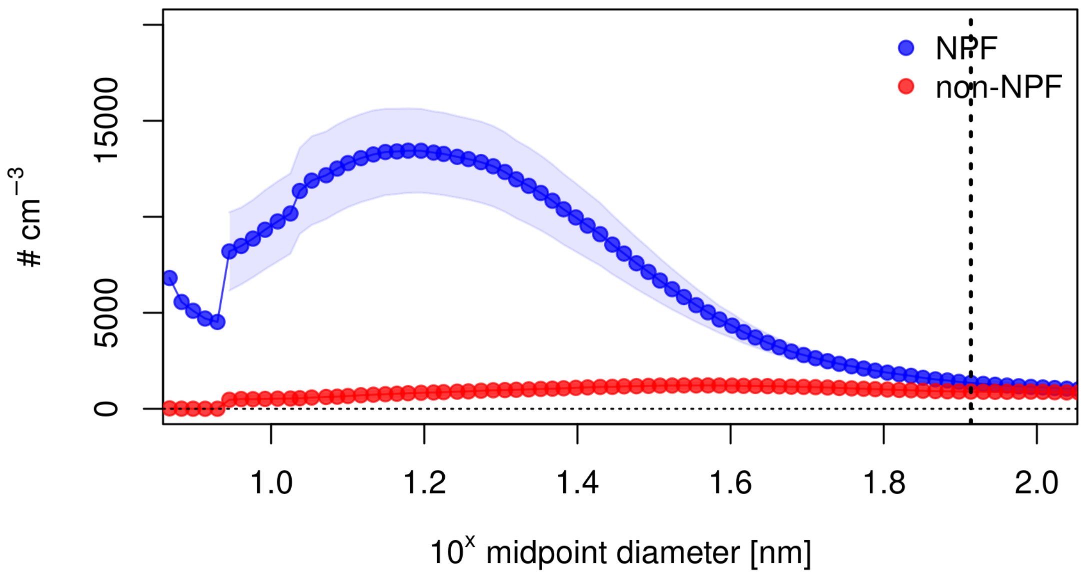
Figure 4. Please provide the log and log-log scales at least on the answer to the reviewer. I would like to see that figure on that scale. The Non-NPF events line is not clearly seen.

The authors have provided copies of the different plots that were generated when crafting the response to reviewer 1. We still believe the normal-normal scale is the best way to observe the difference in figure 4. The figures can be viewed below:

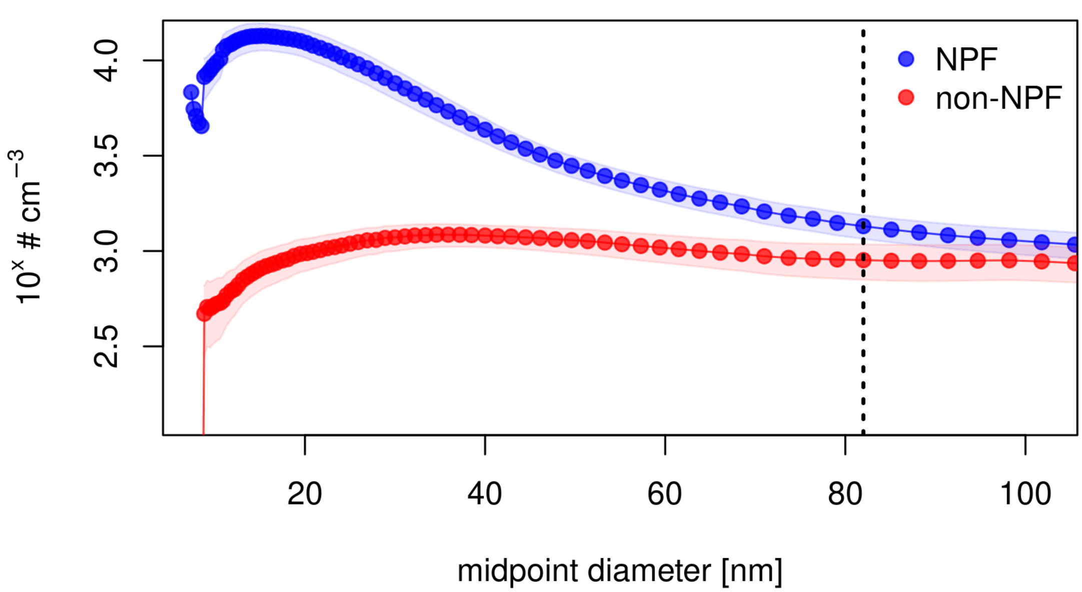
Normal scale (proposed plot):



X axis on log scale:



Y axis on log scale:



Both axis on log scale: