Supporting Information

Gordon et al.

S1. Ion concentrations in the GLOMAP aerosol model

We consider two sources of ions in the atmosphere: radon and galactic cosmic rays. Radon is dominant at the land surface, where most biogenic nucleation is likely to happen. Ion production rates from radon are read in from look-up tables [1]. Above the surface and over the ocean, cosmic ray ionisation is more important. The ionization rates from cosmic rays are calculated from lookup tables [2] which are provided for several solar cycles, so the effect of the Sun's magnetic field can be incorporated via the heliospheric modulation potential. The technique of Fraser-Smith [3] is used to calculate the geomagnetic cut-off rigidity from the International Geomagnetic Reference Field coefficients. These are available with five-yearly time resolution so are interpolated within the five-year periods, then the atmospheric depth (which determines the interaction probability of a cosmic ray) and the heliospheric modulation potential are spatially interpolated across the model grid-boxes.

The small-ion concentration of either sign, $[n_{\pm}] = [n_{+}] = [n_{-}]$, is calculated from the steady state solution of the ion balance equation [4]

$$\mathrm{d}[n_{\pm}]/\mathrm{d}t = q - \alpha[n_{\pm}]^2 - k_i[n_{\pm}] \qquad [1]$$

where q is the ion pair production rate from GCRs and α is the ion-ion recombination coefficient ($\mathrm{cm^3 s^{-1}}$). The factor 2 in Eq. 4 accounts for nucleation from both positive and negative ions. The ion loss rate, k_i, is due to the condensation sink, CS, and ion-induced nucleation, so that $k_i = CS + J_{iin}/2\,[n_{\pm}]$ where $J_{iin}/2\,[n_{\pm}]$ is given by Eq. 4 and the steady state concentration of small ions is $[n_{\pm}] = [(k_i^2 + 4\alpha q)^{0.5} - k_i]/2\alpha$. From Eq. 1, J_{iin} saturates at $2q$ at high nucleation rates (see Ref. [5] Fig. 2).

S2. Simulating the pre-industrial atmosphere

The concentrations of key precursor gases for particle formation are compared between present-day and pre-industrial in Fig. S1. The percentage changes between pre-industrial and present-day are compared in Table S1 for summer and winter in the two hemispheres. The sulphuric acid concentration is substantially higher in the present day atmosphere due to much higher emissions, while the organic concentrations are higher in the pre-industrial atmosphere due to lower sinks.

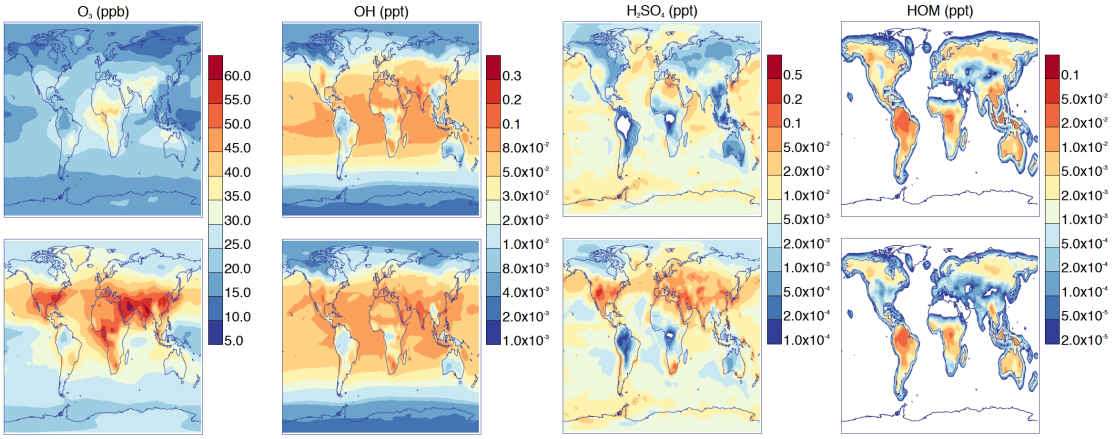

Fig. S1. Concentrations of key gases: ozone, hydroxyl radicals, sulphuric acid and HOMs in pre-industrial (top row) and present-day atmospheres (bottom row) at cloud base level, annually averaged.

S3. Modelled changes in particle concentrations and further discussion of particle numbers in the Amazon region

In Fig. S2 we present the seasonal cycle in surface 3 nm particle concentrations with and without pure biogenic nucleation, and the change when pure biogenic nucleation is included. This figure shows that the strongest effects are in the present-day in summertime in boreal regions, Australia, southern Africa and the Amazon region. Fig. S3 shows the effect of pure biogenic nucleation on cloud-level CCN concentrations in months chosen to reflect the Amazon wet and dry seasons (February and August). In Sect. S5, we further show that the present-day concentrations are in good agreement with observations at a diverse range of surface sites.

As discussed in the main text, on average our model predicts greater numbers of particles in the Amazon than observations suggest. According to the review by Martin *et al* [6], the mean number concentration in the Aitken mode is 239 $\mathrm{cm^{-3}}$ and that in the accumulation mode is 177 $\mathrm{cm^{-3}}$, so the total concentration of particles of at least 70 nm in diameter (N70, a reasonable proxy for CCN, usually equivalent to a supersaturation between 0.4% and 1%) is $\sim 300\,\mathrm{cm^{-3}}$ in the wet season. Our model

Table S1. Relative changes between present-day and pre-industrial atmosphere at cloud level, averaged over the month for January and July, and averaged over the entire year in the "Annual" column. The value quoted is the percentage increase in the mean in the present-day compared to the pre-industrial atmosphere. Pure biogenic nucleation is included.

Quantity	Change w.r.t. pre-industrial (%)				
	Jan. NH	Jul. NH	Jan. SH	Jul. SH	Annual
O_3	76.9	70.2	32.5	37.9	59.1
OH	41.7	21.3	-0.9	-0.1	14.6
H_2SO_4	361.4	84.4	0.8	42.3	79.3
HOM	-69.4	-36.1	-21.6	-26.4	-39.8
total J	1255.0	445.3	23.0	341.5	491.6
J_R	1264.5	164.3	24.5	343.2	484.0
J_{org}	-90.3	-47.5	-33.6	-65.3	-57.8
N_3	136.4	36.8	3.0	10.6	42.9
N_{70}	117.0	36.9	6.0	7.5	41.8
CCN 1%	122.2	42.2	4.6	7.3	44.4
CCN 0.2%	99.3	72.0	7.5	9.5	54.2

predicts N70 of $\sim 500 - 800\,\mathrm{cm}^{-3}$ in the wet season (higher near Manaus, lower near the coast, see Fig. S3) so it is still high, but within a factor 2 or 3, which may be larger than the measurement uncertainty but is certainly within our best estimate of the model parametric uncertainty [7]. In February in Manaus, we predict CCN concentrations at 0.2% supersaturation to be $127\,\mathrm{cm}^{-3}$ without pure biogenic nucleation, and pure biogenic nucleation increases this to $299\,\mathrm{cm}^{-3}$. Observations in the wet season in the pristine forest near Manaus are considerably lower, at $30 - 80\,\mathrm{cm}^{-3}$. However, our model averages over both the pristine forest and the Manaus pollution plume in this area so would be expected to yield higher concentrations than the pristine observations. There is the additional challenge that Manaus is so close to the Equator that it is close to the boundary between the wet season and the dry season. The observed transition season concentrations of $200 - 300\,\mathrm{cm}^{-3}$ may be more appropriate, which would agree with our model. In the dry season, our model is in reasonably good agreement with observations, with on average $400 - 800\,\mathrm{cm}^{-3}$ CCN 0.2% (Fig. S3, Ref. [6]).

Overprediction of nucleation rates in the Amazon region suggests that pure biogenic nucleation may be suppressed there. The implications of the CLOUD results for the global atmosphere thus depend on whether a mechanism that suppresses pure biogenic nucleation exists, and if it does, whether or not it has a seasonal dependence, or is localised to regions with similar characteristics to the Amazon. For example, it could be associated with high humidity which increases the effective condensation sink [8, 9], high isoprene concentrations [10], high temperatures, or high levels of peroxy radicals, or it could be present globally all year round.

We note that overprediction of CCN (even without pure biogenic nucleation) is a feature of many global aerosol models. The Aerocom assessment of 15 global aerosol models [11] shows annual mean N100 concentrations in the Amazon of over $500\,\mathrm{cm}^{-3}$ and N30 concentrations of closer to $1000\,\mathrm{cm}^{-3}$. This suggests that there are general model weaknesses in the Amazon which makes it difficult to say anything with confidence about the effects an additional aerosol source would have on this region.

In Ref. [6], it is observed that particles in the nucleation and Aitken modes in the wet season rarely grow to larger sizes (although it is certainly the case that some condensational growth does occur). This is *a priori* surprising since the terpene and condensable organic concentrations are very high in this area. The most likely explanation must be very high losses: high precipitation frequency in the wet season and high condensation sink in the dry season, or very strong vertical mixing that prevents particle growth being observed adequately from a single surface location. Any losses are also likely to be inhomogenously distributed, and therefore their non-linear effects could be easily underestimated in a low-resolution model like ours.

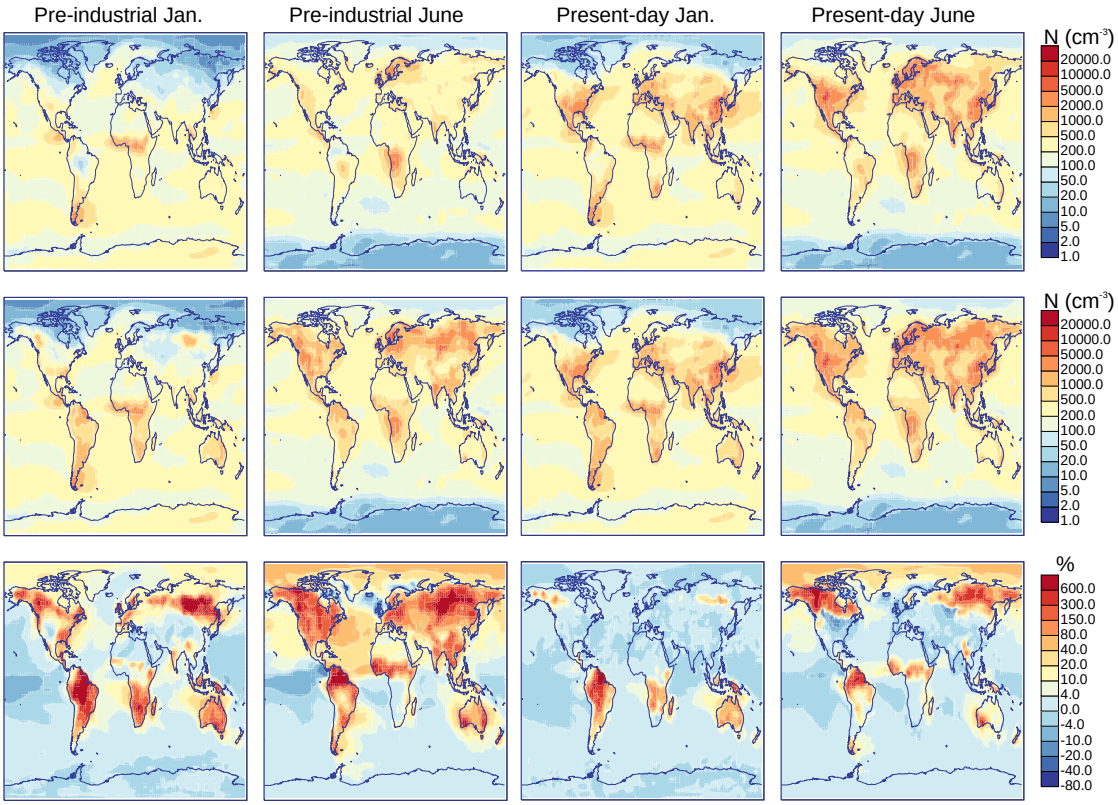

Fig. S2. Monthly average surface level concentrations of particles larger than $3\ \mathrm{nm}$ in diameter, along each row pre-industrial January, pre-industrial June, present-day January and present-day June. Top: without pure biogenic nucleation. Middle: with pure biogenic nucleation. Bottom: percentage changes when pure biogenic nucleation is added.

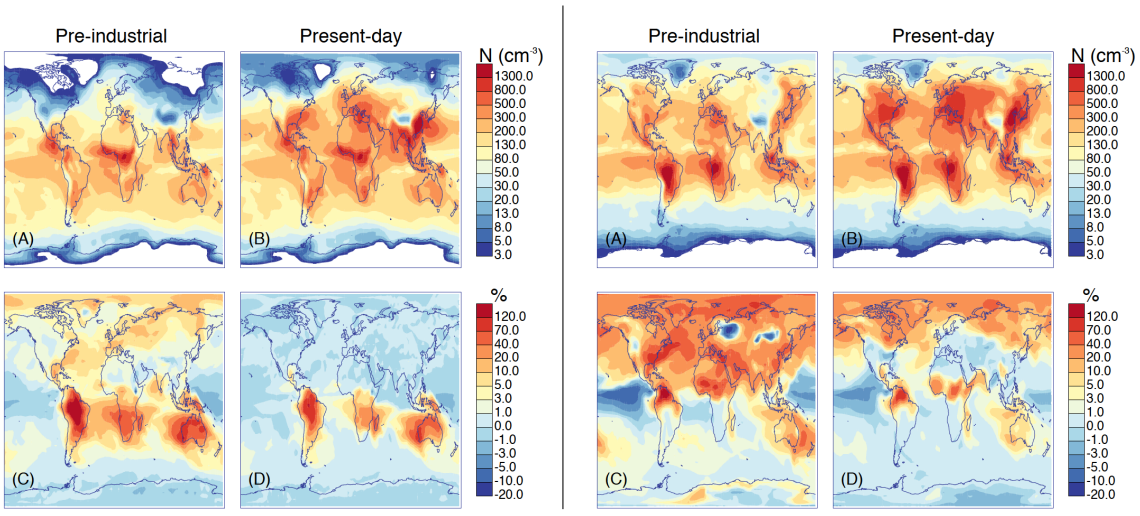

Fig. S3. Concentrations of cloud condensation nuclei calculated at 0.2% supersaturation, in cm^{-3}, in February (left four panels) and August (right four panels). On either side of the line, average CCN concentrations at cloud base level over the month in (A) pre-industrial and (B) present-day conditions are shown, and, below these in subfigures (C, D), the percentage changes to these concentrations when pure biogenic nucleation is introduced.

S4. Diurnal cycles of particle formation rates

In Fig. S4, we show model predictions of the diurnal cycle of particle formation in July at Pallas and at the most studied field site, Hyytiälä, also in Finland. We predict that pure biogenic nucleation contributes significantly to the nucleation rate at both sites in July. One would expect APi-TOF data at Hyytiälä in July to show clusters of HOMs both with and without sulphuric acid. However, identifying the absence of sulphuric acid from clusters large enough to be equivalent to nucleated particles in mass spectra from field measurements at Hyytiälä has not been possible [12]. Furthermore, nucleation measurements at Hyytiälä are usually made in spring, when pure biogenic nucleation is predicted to make a much smaller contribution (see also Fig. S8, below). At Pallas, on the other hand, nucleation events will be rarer and signals are likely to be smaller, but the background particle concentrations responsible for the condensation sink are lower. Therefore, the peaks in nucleation at 200 and 520 hours into the month in Fig. S4 might well lead to observable 'banana'-type events, and our model suggests that all nucleation at this site should be dominated by pure biogenic processes. We note that while the pure biogenic nucleation mechanism can in principle operate both day and night, the model does not predict any nucleation at night, principally because terpene emissions are higher during daytime.

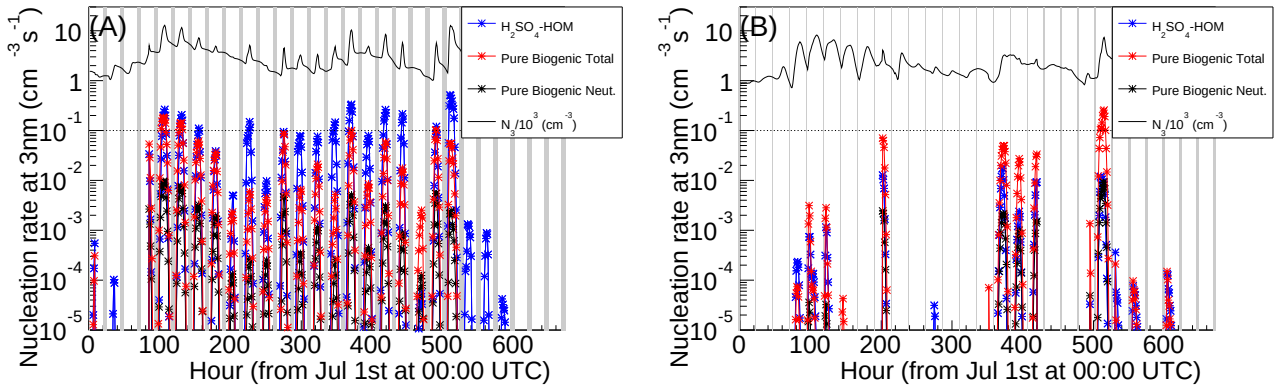

Fig. S4. Modelled diurnal cycles of nucleation rates and particle concentrations at present-day (A) Hyytiälä (61.85°N, 24.28°E) and (B) Pallas (68.00°N, 24.23°E) in the first four weeks of July 2008. July is the month where the pure biogenic nucleation rate at Hyytiälä is strongest. Intervals between sunset and sunrise are marked in grey. The wind changes direction around 90 hours into the month. Observable nucleation events are likely when the nucleation rate is above around $0.1\,\mathrm{cm^{-3}s^{-1}}$, indicated by the dotted line. At Pallas, Aitken mode particles are transported from nucleation happening elsewhere, which explains the daytime peaks in particle number concentration even when nucleation rates at Pallas are very low.

We also show the diurnal cycle in February (wet season) and August (dry season) at the most studied observation site in the Amazon, Manacapuru, in Fig. S5. Manacapuru is in the same model gridbox as the Amazon Tall Tower Observatory and as Manaus. At the surface level, approximately 0 to 30 m above ground level, we predict quite a substantial amount of nucleation, but rarely, perhaps never, enough to produce observable banana-type events, especially in the dry season.

Further, we show the evolution of the size distributions at the surface level at Hyytiala and Pallas in August in Fig. S6, and in Manacapuru for February and August in Fig. S7. Clear nucleation events are seen at both Finnish sites which are similar to the observations detailed in, for example, Refs. [13] or [14] for Hyytiala and [15] or [16] for Pallas. The size distributions also show that in the dry season in the Amazon essentially no nucleation is predicted, while in the wet season very weak additions to the Aitken mode are predicted. We speculate that these signals would rarely, if ever, be observable as nucleation events due to the complicated and inhomogenous meteorology.

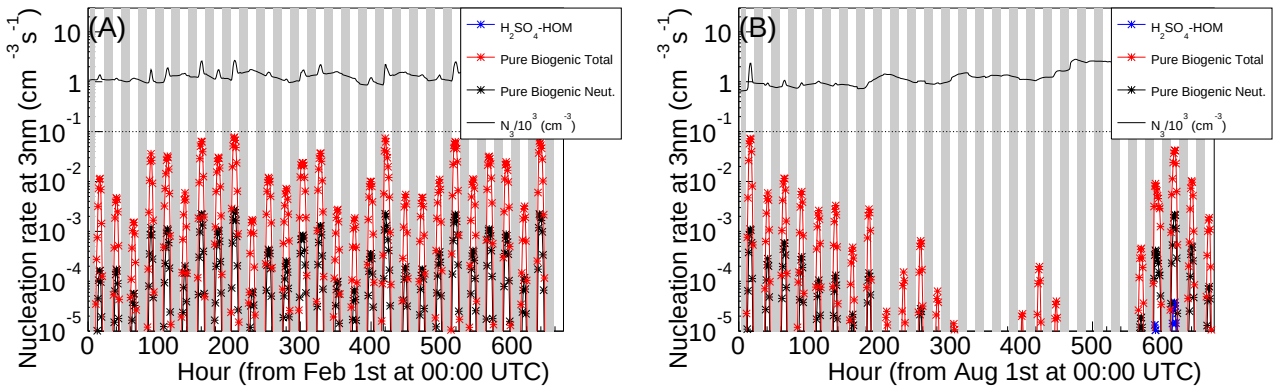

Fig. S5. Modelled diurnal cycles of nucleation rates at present-day Manacapuru (3.30°S, 60.62°W) in (A) the wet season (February) and (B) the dry season (August), in 2008. The particle number concentration shows small spikes during nucleation events, often increasing from around $1000\,\mathrm{cm}^{-3}$ to around $2000\,\mathrm{cm}^{-3}$ due to nucleation. However, in these figures, our spatial model resolution smears out much larger, more local fluctuations in the particle concentrations (which can easily be a factor 10), and so in reality such small spikes would be difficult, and probably impossible, to discern in atmospheric observations.

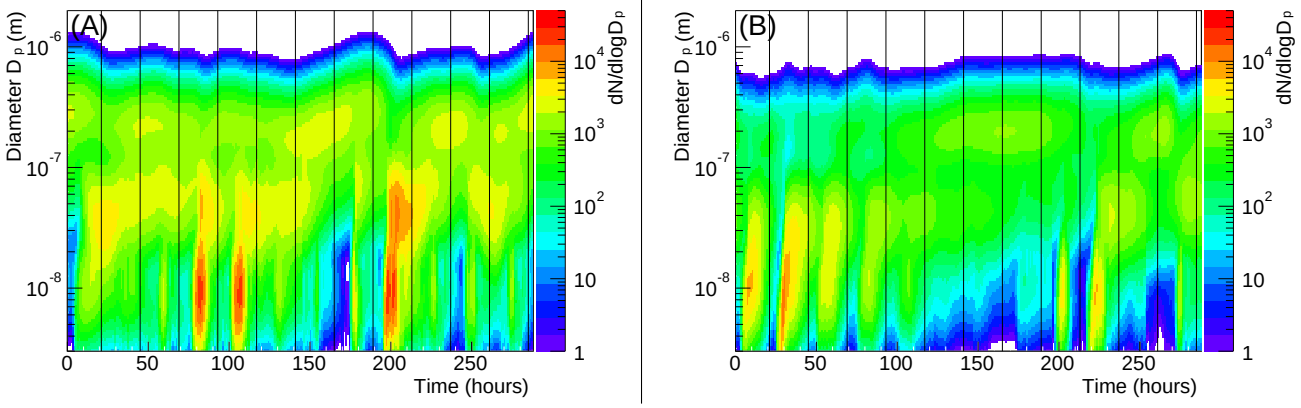

Fig. S6. Modelled diurnal cycles of particle size distribution $dN/d\log D_p$ at present-day (A) Hyytiala and (B) Pallas, in the first twelve days of August (UTC time). The vertical lines mark midnight Finnish local time (UTC+3). In this figure $d\log D_p = 0.02$.

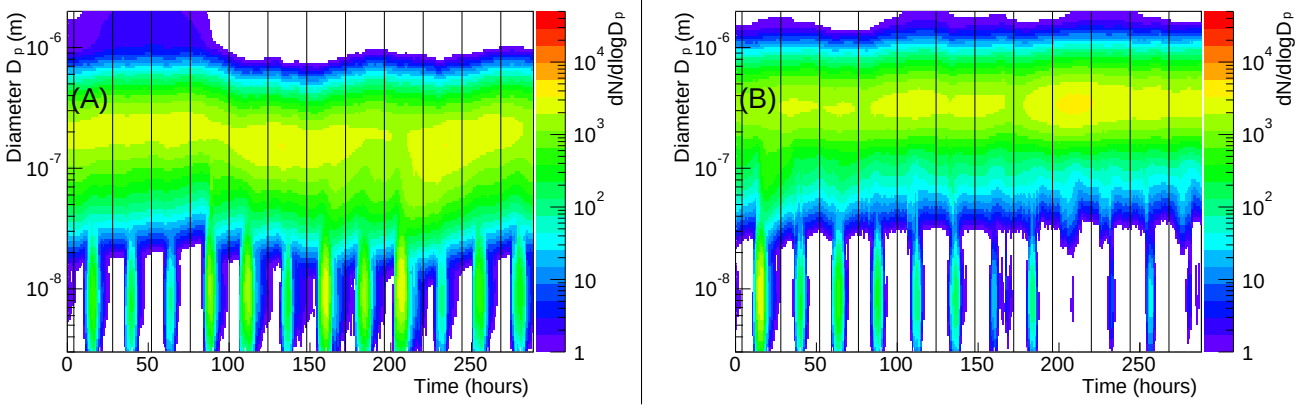

Fig. S7. Modelled diurnal cycles of particle size distribution at present-day Manacapuru in (A) the wet season (the first twelve days of February, UTC time) and (B) the dry season (the first twelve days of August). Like the fluctuations in particle number concentration shown in Fig. S5, the small spikes in the nucleation mode would be difficult to see in observation data. The vertical lines mark midnight local time (UTC-4). In this figure $d\log D_p = 0.02$.

S5. Model evaluation against measurements

The model was evaluated by comparing the particle number concentrations it predicts to those measured at 37 surface sites (Fig. S8). The first thirty-six are those used in Ref. [17]. We also added previously unpublished data recorded in 2010, 2011 and 2013 from a condensation particle counter with a 4 nm cut-off diameter at the East Trout Lake Global Atmosphere Watch station (54.35° N 104.98° W) because our model predicts a significant contribution from pure biogenic nucleation in central Canada. When measurements from multiple years are available, the data from the months in each year were averaged. Averaging over all sites and over the whole year, we find including pure biogenic nucleation leads to a modest improvement in the model bias from -42% to -41%. In summer the bias changes from -36% to -34% when pure biogenic nucleation is included and in winter it is unchanged at -53%. The overall low bias, particularly in winter, is likely to be because we do not include the effects of ammonia or anthropogenic organic molecules on nucleation in our model. Fig. S8 shows that pure biogenic nucleation strongly affects particle concentrations only at East Trout Lake, Listvyanka and Point Barrow.

We also compared the daily mean particle number concentrations from the model to observations made during the ARCTAS campaign [18] in spring and summer 2008 (Fig. S9). We interpolate these modelled particle concentrations within model grid boxes to match the locations of one-minute-averaged condensation particle counter measurements from a NASA P3-B aircraft. We note that the campaign during summer, when pure biogenic nucleation has most effect, was designed to investigate the influence of boreal forest fires. Our low model resolution and averaged fire emissions inventory mean that close agreement between the model and the measurements is not expected. However, we attempt the comparison anyway as the measurements are, unusually, in a particularly relevant region where pure biogenic nucleation is predicted to have a strong effect. Averaged over all altitudes, the model without pure biogenic nucleation is biased low (−57.9%); including pure biogenic nucleation leads to a smaller bias of −37.4%. Most of the pure biogenic contribution is in a band of latitudes from 51 to 62°N, corresponding to flights starting from Cold Lake between 26 June and 14 July 2008. While the relatively large underprediction of particle number without pure biogenic nucleation suggests a particle formation pathway for pristine regions like ours may be needed, large local deviations of the model and measurements evident in Fig. S9 preclude any firm conclusion being drawn. These deviations could be attributed to the fire emissions or low model resolution discussed earlier, temperature or chemistry effects on new particle formation that we did not account for, or uncertainties in the observation data.

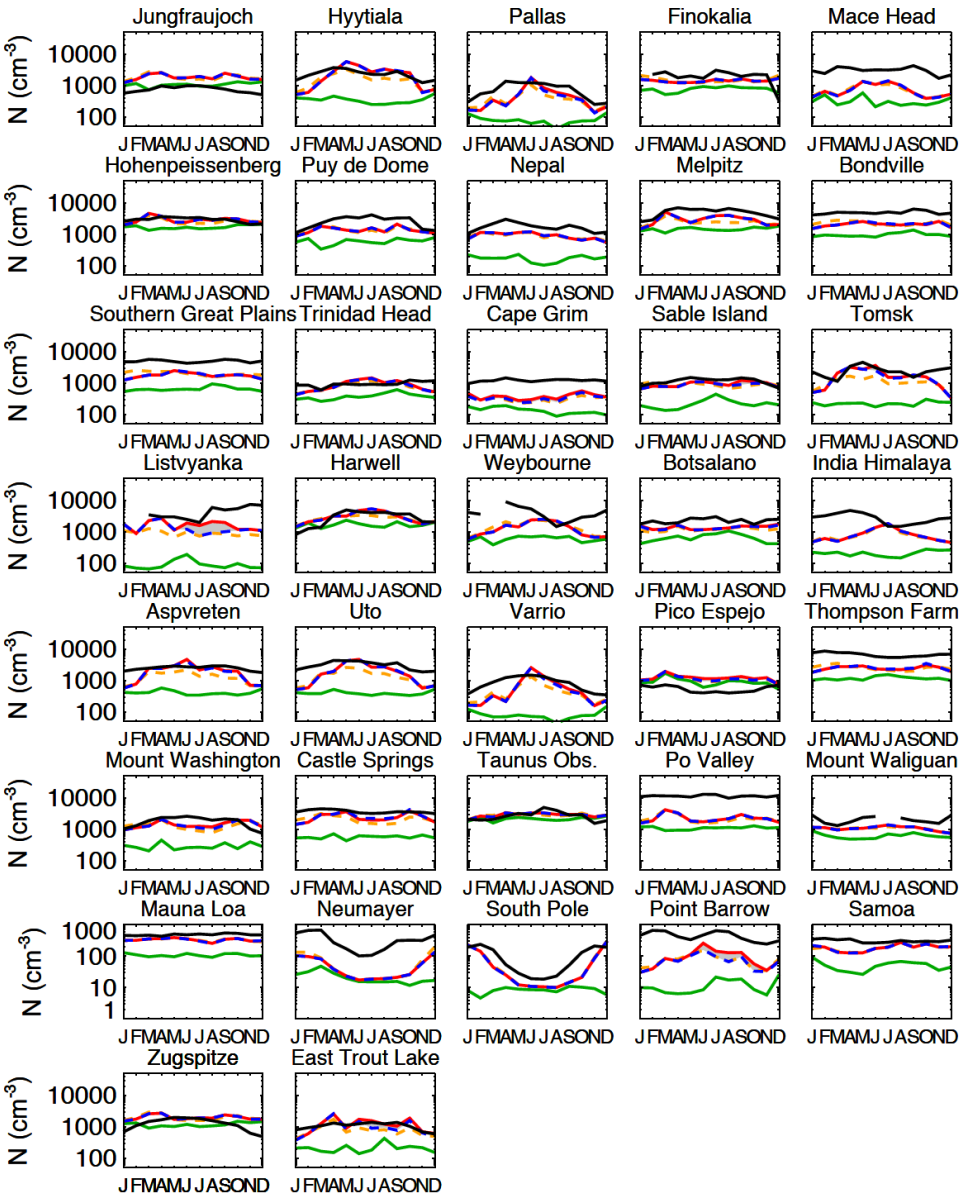

Fig. S8. Particle concentrations at selected measurement sites [17, 19], in black, measured by counters with cut off sizes varying from $3\,\mathrm{nm}$ to $14\,\mathrm{nm}$, compared to model predictions.The red curve shows the particle concentrations predicted by the baseline nucleation mechanisms, numbered 1-3 in the main text and including pure biogenic nucleation. The blue dotted curve shows the particle concentrations predicted without pure biogenic nucleation. A grey band is drawn between these two curves. The orange dotted curve shows particle concentrations predicted by the parametrisation of Paasonen et al [20], including a component of pure biogenic nucleation proportional to the square of the organic concentration. The dark green curve shows particles from primary emissions only.

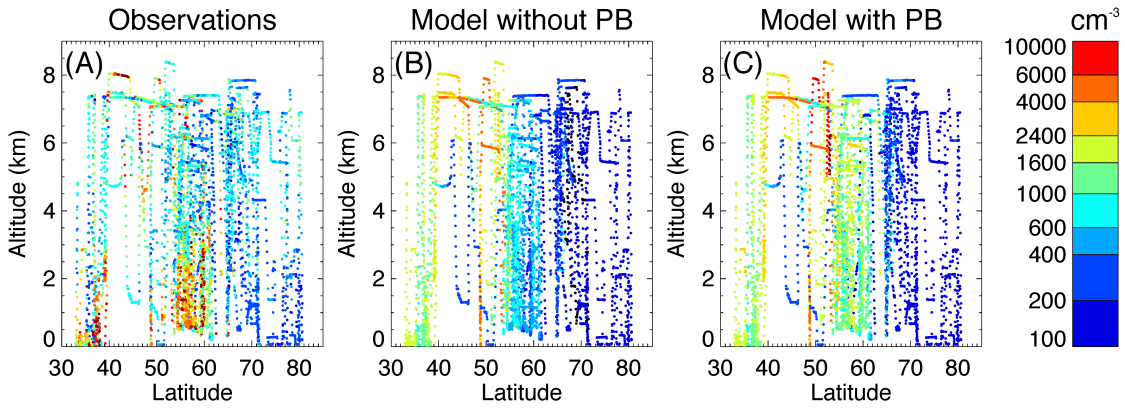

Fig. S9. Vertical profiles of particle number concentrations calculated at S.T.P. (all particles of at least 3 nm in diameter), (A) measured during the ARCTAS campaign [18] in 2008, (B) modelled without pure biogenic nucleation (PB) and (C) modelled including pure biogenic nucleation.

S6. Further discussion of uncertainties and summary tables of sensitivity studies

Table S2 gives the changes to CCN concentrations due to pure biogenic nucleation in different scenarios (e.g. if organic nucleation is temperature-dependent, or if primary emissions are higher in our simulation than in reality). Following Ref. [7], the high primary emissions scenario corresponds to a doubling of biomass burning primary particulate emissions, a reduction in the mode diameter of biomass burning primary emissions from 150 nm to 100 nm, and a factor 2.5 increase in the sea spray flux. The low emissions scenario corresponds to a halving of biomass burning emissions, an increase in the mode diameter to 175 nm and a reduction in the sea spray flux by 60%. Table S3 gives the corresponding changes to radiative forcing.

Table S2. Annual average global mean changes to the concentrations of particles larger than 70 nm in diameter, a proxy for CCN, when pure biogenic nucleation is introduced. The differences between the contents of the first row and the subsequent rows give an indication of the sensitivity of the analysis to different perturbations. The temperature dependence is added for both the baseline organic and the pure biogenic nucleation mechanisms, so changes both the control and perturbed simulations.

Variation	Change to CCN 0.2% (PI) (%)	Change to CCN 0.2% (PD) (%)
Add pure biogenic nucleation (PBN)	12	4
Add PBN with trebled yield	19	6
Add PBN with one-third yield	4	1
Add T dependence	7	2
Double pre-industrial volcanic SO2	11	-
Halve pre-industrial volcanic SO2	13	-
High primary emissions	7	2
Low primary emissions	14	5
Use baseline J from Ref. [20]	14	5

Table S3. Annual average global mean radiative forcings (including pure biogenic nucleation) and absolute changes to forcing when pure biogenic nucleation is introduced.

Variation	Forcing (PD-PI) Wm^{-2}	Change to forcing with pure biogenic nucleation (Wm^{-2})
Add pure biogenic nucleation (PBN)	-0.60	+0.22
Add PBN with trebled yield	-0.52	+0.30
Add PBN with one-third yield	-0.72	+0.10
Add T dependence	-0.64	+0.14
High primary emissions	-0.63	+0.17
Low primary emissions	-0.54	+0.33
Use baseline J from Ref. [20]	-0.63	+0.24
Exclude area close to Equator	-0.62	+0.20

In the main text, we discussed various sources of uncertainty, such as uncertainties in the CLOUD experimental measurements or HOM yields, in the possible temperature dependence of pure biogenic nucleation, and in primary emissions. The experiment in which we determine the sensitivity to primary emissions also tests our sensitivity to uncertainty in the condensation sink.

In the particularly interesting Amazon region, this uncertainty will be strongly influenced by the quantity and mode diameter of biomass burning emissions [7]. The 150 nm diameter used, following Ref. [21], is already larger than that of fresh smoke particles (averaged over all vegetation types) of 117 nm [22], and in theory condensation of secondary organic vapours in our model should increase this diameter to the aged diameter of 235 nm in a few days. However, the aged diameter in reality depends on cloud processing of organics as well as condensation, and this is not included in our model. To compensate for this, the larger fresh diameter of 150 nm is used by default in GLOMAP, but it has a large uncertainty [7]. We increase it to 175 nm in our sensitivity study with high primary emissions, which is still well within the uncertainty and the ranges for different phases of typical Amazon burning found in Ref. [23].

Another possible source of uncertainty in the Amazon region, which we are not able to quantify directly, is the condensation sink during periods of high humidity [8, 9] or aerosol-cloud interaction [24]. Clouds in the Amazon region are likely to be strongly affected by both of these. For the latter, impaction scavenging in GLOMAP is due to raindrops but the only loss mechanism to cloud droplets is nucleation scavenging. In a cloud, the coagulation sink should surely dramatically increase. This is not modelled. The effective condensation sink in a monodisperse cloud with 100 droplets cm^{-3} of diameter 15.6 μm (corresponds to LWC 0.2 gm^{-3}) is 0.16 s^{-1}. We investigated this further by including a crude treatment of this effect in our model. The perturbation to the model results in this test was small, mostly because clouds rarely cover more than half a model gridbox in areas where pure biogenic nucleation is important, and therefore the average effect is smeared out. With higher model resolution, however, the effect would likely be stronger due to the nonlinear nature of nucleation and survival probability.

1. Zhang, K et al. (2011) Radon activity in the lower troposphere and its impact on ionization rate: a global estimate using different radon emissions. *Atmospheric Chemistry and Physics* **11**, 7817–7838.
2. Usoskin, I. G, Kovaltsov, G. A, & Mironova, I. A. (2010) Cosmic ray induced ionization model CRAC:CRII: An extension to the upper atmosphere. *Journal of Geophysical Research* **115**, 6.
3. Fraser-Smith, A. C. (1987) Centered and eccentric geomagnetic dipoles and their poles, 1600–1985. *Reviews of Geophysics* **25**, 1–16.
4. Franchin, A et al. (2015) Experimental investigation of ion-ion recombination under atmospheric conditions. *Atmospheric Chemistry and Physics* **15**, 7203–7216.
5. Kirkby, J et al. (2016) Ion-induced nucleation of pure biogenic particles. *Nature* **533**, 521–526.
6. Martin, S et al. (2010) Sources and properties of Amazonian aerosol particles. *Reviews of Geophysics* **48**.
7. Lee, L. A et al. (2013) The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei. *Atmospheric Chemistry and Physics* **13**, 8879–8914.
8. Rose, C et al. (2015) Frequent nucleation events at the high altitude station of Chacaltaya (5240 m a.s.l.), Bolivia. *Atmospheric Environment* **102**, 18 – 29.
9. Falvey, M & Garreaud, R. D. (2005) Moisture variability over the South American Altiplano during the South American low level jet experiment (SALLJEX) observing season. *Journal of Geophysical Research: Atmospheres* **110**.
10. Kiendler-Scharr, A et al. (2009) New particle formation in forests inhibited by isoprene emissions. *Nature* **461**, 381–384.
11. Mann, G. W et al. (2014) Intercomparison and evaluation of global aerosol microphysical properties among AeroCom models of a range of complexity. *Atmospheric chemistry and physics* **14**, 4679–4713.
12. Schobesberger, S et al. (2013) Molecular understanding of atmospheric particle formation from sulfuric acid and large oxidized organic molecules. *Proceedings of the National Academy of Sciences* **110**, 17223–17228.
13. Dal Maso, M et al. (2005) Formation and growth of fresh atmospheric aerosols: eight years of aerosol size distribution data from SMEAR II, Hyytiala, Finland. *Boreal Environment Research* **10**, 323.
14. Kulmala, M et al. (2012) Measurement of the nucleation of atmospheric aerosol particles. *Nature Protocols* **7**, 1651–1667.
15. Väänänen, R et al. (2013) Analysis of particle size distribution changes between three measurement sites in northern Scandinavia. *Atmos. Chem. Phys* **13**, 11887–11903.
16. Komppula, M, Sihto, S.-L, Korhonen, H, Lihavainen, H, Kerminen, V.-M, Kulmala, M, & Viisanen, Y. (2006) New particle formation in air mass transported between two measurement sites in northern finland. *Atmospheric Chemistry and Physics* **6**, 2811–2824.
17. Spracklen, D. V et al. (2010) Explaining global surface aerosol number concentrations in terms of primary emissions and particle formation. *Atmospheric Chemistry and Physics* **10**, 4775–4793.
18. Jacob, D. J et al. (2010) The Arctic Research of the Composition of the Troposphere from Aircraft and Satellites (ARCTAS) mission: design, execution, and first results. *Atmospheric Chemistry and Physics* **10**, 5191–5212.
19. Torseth, K et al. (2012) Introduction to the European Monitoring and Evaluation Programme (EMEP) and observed atmospheric composition change during 1972-2009. *Atmospheric Chemistry and Physics* **12**, 5447–5481.
20. Paasonen, P et al. (2010) On the roles of sulphuric acid and low-volatility organic vapours in the initial steps of atmospheric new particle formation. *Atmospheric Chemistry and Physics* **10**, 11223–11242.
21. Stier, P et al. (2005) The aerosol-climate model ECHAM5-HAM. *Atmospheric Chemistry and Physics* **5**, 1125–1156.
22. Janhäll, S, Andreae, M. O, & Pöschl, U. (2010) Biomass burning aerosol emissions from vegetation fires: particle number and mass emission factors and size distributions. *Atmospheric Chemistry and Physics* **10**, 1427–1439.
23. Costa, M. A. M et al. (2012) Real-time sampling of particulate matter smaller than 2.5 μm from Amazon forest biomass combustion. *Atmospheric environment* **54**, 480–489.
24. Pierce, J. R, Croft, B, Kodros, J. K, D'Andrea, S. D, & Martin, R. V. (2015) The importance of interstitial particle scavenging by cloud droplets in shaping the remote aerosol size distribution and global aerosol-climate effects. *Atmospheric Chemistry and Physics* **15**, 6147–6158.