Aerosol

From Wikipedia, the free encyclopedia

This article is about the physical properties of an aerosol. For pressurized cans designed to produce an aerosol, see Aerosol spray.

Not to be confused with aerosil.

Mist and clouds are aerosols.

An **aerosol** is a suspension of fine solid particles or liquid droplets in air or another gas.[1] Aerosols can be natural or anthropogenic. Examples of natural aerosols are fog, dust, forest exudates and geyser steam. Examples of anthropogenic aerosols are haze, particulate air pollutants and smoke.[dubious – discuss][1] The liquid or solid particles have diameters typically <1 μm; larger particles with a significant settling speed make the mixture a suspension, but the distinction is not clear-cut. In general conversation, *aerosol* usually refers to an aerosol spray that delivers a consumer product from a can or similar container. Other technological applications of aerosols include dispersal of pesticides, medical treatment of respiratory illnesses, and combustion technology.[2] Diseases can also spread by means of small droplets in the breath, also called aerosols (or sometimes bioaerosols).[3]

Aerosol science covers generation and removal of aerosols, technological application of aerosols, effects of aerosols on the environment and people, and other topics.[1]

Contents

Definitions

Photomicrograph made with a Scanning Electron Microscope (SEM): Fly ash particles at 2,000× magnification. Most of the particles in this aerosol are nearly spherical.

An aerosol is defined as a suspension system of solid or liquid particles in a gas. An aerosol includes both the particles and the suspending gas, which is usually air.[1] Frederick G. Donnan presumably first used the term *aerosol* during World War I to describe an aero-solution, clouds of microscopic particles in air. This

term developed analogously to the term hydrosol, a colloid system with water as the dispersed medium.[4] *Primary aerosols* contain particles introduced directly into the gas; *secondary aerosols* form through gas-to-particle conversion.[5]

Various types of aerosol, classified according to physical form and how they were generated, include dust, fume, mist, smoke and fog.[6]

There are several measures of aerosol concentration. Environmental science and health often uses the *mass concentration* (*M*), defined as the mass of particulate matter per unit volume with units such as µg/m^3. Also commonly used is the *number concentration* (*N*), the number of particles per unit volume with units such as number/m^3 or number/cm^3.[7]

The size of particles has a major influence on their properties, and the aerosol particle radius or diameter (d_p) is a key property used to characterise aerosols.

Aerosols vary in their dispersity. A *monodisperse* aerosol, producible in the laboratory, contains particles of uniform size. Most aerosols, however, as *polydisperse* colloidal systems, exhibit a range of particle sizes.[8] Liquid droplets are almost always nearly spherical, but scientists use an *equivalent diameter* to characterize the properties of various shapes of solid particles, some very irregular. The equivalent diameter is the diameter of a spherical particle with the same value of some physical property as the irregular particle.[9] The *equivalent volume diameter* (d_e) is defined as the diameter of a sphere of the same volume as that of the irregular particle.[10] Also commonly used is the aerodynamic diameter.

Size distribution

The same hypothetical log-normal aerosol distribution plotted, from top to bottom, as a number vs. diameter distribution, a surface area vs. diameter distribution, and a volume vs. diameter distribution. Typical mode names are shows at the top. Each distribution is normalized so that the total area is 1000.

For a monodisperse aerosol, a single number—the particle diameter—suffices to describe the size of the particles. However, more complicated particle-size distributions describe the sizes of the particles in a polydisperse aerosol. This distribution defines the relative amounts of particles, sorted according to size.[11] One approach to defining the particle size distribution uses a list of the sizes of every particle in a sample. However, this approach proves tedious to ascertain in aerosols with millions of particles and awkward to use. Another approach splits the complete size range into intervals and finds the number (or proportion) of particles in each interval. One then can visualize these data in a histogram with the area of each bar representing the proportion of particles in that size bin, usually normalised by dividing the number of particles in a bin by the width of the interval so that the area of each bar is proportionate to the number of particles in the size range that it represents.[12] If the width of the bins tends to zero, one gets the frequency function:[13]

One also usefully can approximate the particle size distribution using a mathematical function. The normal distribution usually does not suitably describe particle size distributions in aerosols because of the skewness associated a long tail of larger particles. Also for a quantity that varies over a large range, as many aerosol sizes do, the width of the distribution implies negative particles sizes, clearly not physically realistic. However, the normal distribution can be suitable for some aerosols, such as test aerosols, certain pollen grains and spores.[16]

A more widely chosen log-normal distribution gives the number frequency as:[16]

The log-normal distribution has no negative values, can cover a wide range of values, and fits many observed size distributions reasonably well.[17]

Other distributions sometimes used to characterise particle size include: the Rosin-Rammler distribution, applied to coarsely dispersed dusts and sprays; the Nukiyama-Tanasawa distribution, for sprays of extremely broad size ranges; the power function distribution, occasionally applied to atmospheric aerosols; the exponential distribution, applied to powdered materials; and for cloud droplets, the Khrgian-Mazin distribution.[18]

Physics

Terminal velocity of a particle in a fluid

For low values of the [Reynolds number](#) (<1), true for most aerosol motion, [Stokes' law](#) describes the force of resistance on a solid spherical particle in a fluid. However, Stokes' law is only valid when the velocity of the gas at the surface of the particle is zero. For small particles (< 1 μm) that characterize aerosols, however, this assumption fails. To account for this failure, one can introduce the [Cunningham correction factor](#), always greater than 1. Including this factor, one finds the relation between the resisting force on a particle and its velocity:[19]

where

        is the resisting force on a spherical particle

        is the dynamic [viscosity](#) of the gas

        is the particle velocity

        is the Cunningham correction factor.

This allows us to calculate the [terminal velocity](#) of a particle undergoing gravitational settling in still air. Neglecting [buoyancy](#) effects, we find:[20]

where

        is the terminal settling velocity of the particle.

The terminal velocity can also be derived for other kinds of forces. If Stokes' law holds, then the resistance to motion is directly proportional to speed. The constant of proportionality is the mechanical mobility (B) of a particle:[21]

A particle traveling at any reasonable initial velocity approaches its terminal velocity [exponentially](#) with an e-folding time equal to the relaxation time:[22]

where:

        is the particle speed at time t

        is the final particle speed

        is the initial particle speed

To account for the effect of the shape of non-spherical particles, a correction factor known as the *dynamic shape factor* is applied to Stokes' law. It is defined as the ratio of the resistive force of the irregular particle to that of a spherical particle with the same volume and velocity:[23]

where:

        is the dynamic shape factor

Aerodynamic diameter

The aerodynamic diameter of an irregular particle is defined as the diameter of the spherical particle with a density of 1000 kg/m^3 and the same settling velocity as the irregular particle.[24]

Neglecting the slip correction, the particle settles at the terminal velocity proportional to the square of the aerodynamic diameter, d_a:[24]

where

= standard particle density (1000 kg/m^3).

This equation gives the aerodynamic diameter:[25]

One can apply the aerodynamic diameter to particulate pollutants or to inhaled drugs to predict where in the respiratory tract such particles deposit. Pharmaceutical companies typically use aerodynamic diameter, not geometric diameter, to characterize particles in inhalable drugs.[*citation needed*]

Dynamics

The previous discussion focussed on single aerosol particles. In contrast, *aerosol dynamics* explains the evolution of complete aerosol populations. The concentrations of particles will change over time as a result of many processes. External processes that move particles outside a volume of gas under study include diffusion, gravitational settling, and electric charges and other external forces that cause particle migration. A second set of processes internal to a given volume of gas include particle formation (nucleation), evaporation, chemical reaction, and coagulation.[26]

A differential equation called the *Aerosol General Dynamic Equation* (GDE) characterizes the evolution of the number density of particles in an aerosol due to these processes.[26]

Change in time = Convective transport + brownian diffusion + gas-particle interactions + coagulation + migration by external forces

Where:

is number density of particles of size category

is the particle velocity

is the particle Stokes-Einstein diffusivity

is the particle velocity associated with an external force

Coagulation

As particles and droplets in an aerosol collide with one another, they may undergo coalescence or aggregation. This process leads to a change in the aerosol particle-size distribution, with the mode increasing in diameter as total number of particles decreases.[27] On occasion, particles may shatter apart into numerous smaller particles; however, this process usually occurs primarily in particles too large for consideration as aerosols.

Dynamics regimes

The Knudsen number of the particle define three different dynamical regimes that govern the behaviour of an aerosol:

where is the mean free path of the suspending gas and is the diameter of the particle.[28] For particles in the *free molecular regime*, $K_n \gg 1$; particles small compared to the mean free path of the suspending gas.[29] In this regime, particles interact with the suspending gas through a series of "ballistic" collisions with gas molecules. As such, they behave similarly to gas molecules, tending to follow streamlines and diffusing rapidly through Brownian motion. The mass flux equation in the free molecular regime is:

where a is the particle radius, P_∞ and P_A are the pressures far from the droplet and at the surface of the droplet respectively, k_b is the Boltzmann constant, T is the temperature, C_A is mean thermal velocity and α is mass accommodation coefficient.[*citation needed*] The derivation of this equation assumes constant pressure and constant diffusion coefficient.

Particles are in the *continuum regime* when $K_n \ll 1$.[29] In this regime, the particles are big compared to the mean free path of the suspending gas, meaning that the suspending gas acts as a continuous fluid flowing round the particle.[29] The molecular flux in this regime is:

where a is the radius of the particle A, M_A is the molecular mass of the particle A, D_{AB} is the diffusion coefficient between particles A and B, R is the ideal gas constant, T is the temperature (in absolute units like kelvin), and $P_{A\infty}$ and P_{AS} are the pressures at infinite and at the surface respectively.[*citation needed*]

The *transition regime* contains all the particles in between the free molecular and continuum regimes or $K_n \approx 1$. The forces experienced by a particle are a complex combination of interactions with individual gas molecules and macroscopic interactions. The semi-empirical equation describing mass flux is:

where I_{cont} is the mass flux in the continuum regime.[*citation needed*] This formula is called the Fuchs-Sutugin interpolation formula. These equations do not take into account the heat release effect.

Partitioning

Condensation and evaporation

Aerosol partitioning theory governs condensation on and evaporation from an aerosol surface, respectively. Condensation of mass causes the mode of the particle-size distributions of the aerosol to increase; conversely, evaporation causes the mode to decrease. Nucleation is the process of forming aerosol mass from the condensation of a gaseous precursor, specifically a vapor. Net condensation of the vapor requires supersaturation, a partial pressure greater than its vapor pressure. This can happen for three reasons:[*citation needed*]

1. Lowering the temperature of the system lowers the vapor pressure.
2. Chemical reactions may increase the partial pressure of a gas or lower its vapor pressure.
3. The addition of additional vapor to the system may lower the equilibrium vapor pressure according to Raoult's law.

There are two types of nucleation processes. Gases preferentially condense onto surfaces of pre-existing aerosol particles, known as **heterogeneous nucleation**. This process causes the diameter at the mode of particle-size distribution to increase with constant number concentration.[30] With sufficiently high supersaturation and no suitable surfaces, particles may condense in the absence of a pre-existing surface, known as **homogeneous nucleation**. This results in the addition of very small, rapidly growing particles to the particle-size distribution.[30]

Activation

Water coats particles in an aerosols, making them *activated*, usually in the context of forming a cloud droplet.[*citation needed*] Following the Kelvin equation (based on the curvature of liquid droplets), smaller particles need a higher ambient relative humidity to maintain equilibrium than larger particles do. The following formula gives relative humidity at equilibrium:

where   is the saturation vapor pressure above a particle at equilibrium (around a curved liquid droplet), p_0 is the saturation vapor pressure (flat surface of the same liquid) and S is the saturation ratio.

Kelvin equation for saturation vapor pressure above a curved surface is:

where r_p droplet radius, σ surface tension of droplet, ρ density of liquid, M molar mass, T temperature, and R molar gas constant.

Solution to the general dynamic equation

There are no general solutions to the general dynamic equation (GDE);[31] common methods used to solve the general dynamic equation include:[32]

- Moment method[33]
- Modal/sectional method,[34] and
- Quadrature method of moments[35][36]/Taylor-series expansion method of moments,[37][38] and

- Monte Carlo method.[39]

Generation and applications

People generate aerosols for various purposes, including:

- as test aerosols for calibrating instruments, performing research, and testing sampling equipment and air filters;[40]
- to deliver deodorants, paints, and other consumer products in sprays;[41]
- for dispersal and agricultural application
- for medical treatment of respiratory disease;[42] and
- in fuel injection systems and other combustion technology.[43]

Some devices for generating aerosols are:[2]

- Aerosol spray
- Atomizer nozzle or nebulizer
- Electrospray
- Electronic cigarette
- Vibrating orifice aerosol generator (VOAG)

Stability of generated aerosol particles

Stability of nanoparticle agglomerates is critical for estimating size distribution of aerosolized particles from nano-powders or other sources. At nanotechnology workplaces, workers can be exposed via inhalation to potentially toxic substances during handling and processing of nanomaterials. Nanoparticles in the air often form agglomerates due to attractive inter-particle forces, such as van der Waals force or electrostatic force if the particles are charged. As a result, aerosol particles are usually observed as agglomerates rather than individual particles. For exposure and risk assessments of airborne nanoparticles, it is important to know about the size distribution of aerosols. When inhaled by humans, particles with different diameters are deposited in varied locations of the central and periphery respiratory system. Particles in nanoscale have been shown to penetrate the air-blood barrier in lungs and be translocated into secondary organs in the human body, such as the brain, heart and liver. Therefore, the knowledge on stability of nanoparticle agglomerates is important for predicting the size of aerosol particles, which helps assess the potential risk of them to human bodies.

Different experimental systems have been established to test the stability of airborne particles and their potentials to deagglomerate under various conditions. A comprehensive system recently reported is able to maintain robust aerosolization process and generate aerosols with stable number concentration and mean size from nano-powders.[44] The deagglomeration potential of various airborne nanomaterials can be also studied using critical orifices.[45] In addition, an impact fragmentation device was developed to investigate bonding energies between particles.[46]

A standard deagglomeration testing procedure could be foreseen with the developments of the different types of existing systems. The likeliness of deagglomeration of aerosol particles in occupational settings can be possibly ranked for different nanomaterials if a reference method is available. For this purpose, inter-laboratory comparison of testing results from different setups could be launched in order to explore the influences of system characteristics on properties of generated nanomaterials aerosols.

Detection

Aerosol can either be measured in-situ or with remote sensing techniques.

In situ observations

Some available in situ measurement techniques include:

- Aerosol mass spectrometer (AMS)
- Differential mobility analyzer (DMA)
- Electrical aerosol spectrometer (EAS)
- Aerodynamic particle sizer (APS)
- Aerodynamic aerosol classifier (AAC)
- Wide range particle spectrometer (WPS)
- Micro-Orifice Uniform Deposit Impactor(MOUDI)
- Condensation particle counter (CPC)
- Epiphaniometer
- Electrical low pressure impactor (ELPI)

- [Aerosol particle mass-analyser](#) (APM)
- [Centrifugal Particle Mass Analyser](#) (CPMA)

Remote sensing approach

Remote sensing approaches include:

- [Sun photometer](#)
- [Lidar](#)
- [Imaging spectroscopy](#)

Size selective sampling

Particles can deposit in the [nose](#), [mouth](#), [pharynx](#) and [larynx](#) (the head airways region), deeper within the respiratory tract (from the [trachea](#) to the [terminal bronchioles](#)), or in the [alveolar region](#).[47] The location of deposition of aerosol particles within the respiratory system strongly determines the health effects of exposure to such aerosols.[48] This phenomenon led people to invent aerosol samplers that select a subset of the aerosol particles that reach certain parts of the respiratory system.[49] Examples of these subsets of the particle-size distribution of an aerosol, important in occupational health, include the inhalable, thoracic, and respirable fractions. The fraction that can enter each part of the respiratory system depends on the deposition of particles in the upper parts of the airway.[50] The inhalable fraction of particles, defined as the proportion of particles originally in the air that can enter the nose or mouth, depends on external wind speed and direction and on the particle-size distribution by aerodynamic diameter.[51] The thoracic fraction is the proportion of the particles in ambient aerosol that can reach the thorax or chest region.[52] The respirable fraction is the proportion of particles in the air that can reach the alveolar region.[53] To measure the respirable fraction of particles in air, a pre-collector is used with a sampling filter. The pre-collector excludes particles as the airways remove particles from inhaled air. The sampling filter collects the particles for measurement. It is common to use [cyclonic separation](#) for the pre-collector, but other techniques include impactors, horizontal [elutriators](#), and large pore [membrane filters](#).[54]

Two alternative size-selective criteria, often used in atmospheric monitoring, are PM_{10} and $PM_{2.5}$. PM_{10} is defined by [ISO](#) as *particles which pass through a size-selective inlet with a 50% efficiency cut-off at 10 μm aerodynamic diameter* and $PM_{2.5}$ as *particles which pass through a size-selective inlet with a 50% efficiency cut-off at 2.5 μm aerodynamic diameter*. PM_{10} corresponds to the "thoracic convention" as defined in ISO 7708:1995, Clause 6; $PM_{2.5}$ corresponds to the "high-risk respirable convention" as defined in ISO 7708:1995, 7.1.[55] The [United States Environmental Protection Agency](#) replaced the older standards for particulate matter based on Total Suspended Particulate with another standard based on PM_{10} in 1987[56] and then introduced standards for $PM_{2.5}$ (also known as fine particulate matter) in 1997.[57]

Atmospheric

Aerosol pollution over northern [India](#) and [Bangladesh](#)

Main article: [Atmospheric particulate matter](#)

Several types of atmospheric aerosol have a significant effect on Earth's climate: volcanic, desert dust, sea-salt, that originating from biogenic sources and human-made. Volcanic aerosol forms in the stratosphere after an eruption as droplets of [sulfuric acid](#) that can prevail for up to two years, and reflect sunlight, lowering temperature. Desert dust, mineral particles blown to high altitudes, absorb heat and may be responsible for inhibiting storm cloud formation. Human-made sulfate aerosols, primarily from burning oil and coal, affect the behavior of clouds.[58]

Although all [hydrometeors](#), solid and liquid, can be described as aerosols, a distinction is commonly made between such dispersions (i.e. clouds) containing activated drops and crystals, and aerosol particles. The [atmosphere of Earth](#) contains aerosols of various types and concentrations, including quantities of:

- natural [inorganic](#) materials: fine dust, sea salt, or water droplets
- natural [organic](#) materials: smoke, [pollen](#), [spores](#), or [bacteria](#)
- [anthropogenic](#) products of combustion such as: smoke, [ashes](#) or dusts

Aerosols can be found in urban [ecosystems](#) in various forms, for example:

- Dust
- Cigarette smoke
- Mist from [aerosol spray](#) cans
- [Soot](#) or fumes in car exhaust

The presence of aerosols in the earth's atmosphere can influence its climate, as well as human health.

Effects

- Volcanic eruptions release large amounts of [sulphuric acid](#), [hydrogen sulfide](#) and [hydrochloric acid](#) into the atmosphere. These gases represent aerosols and eventually return to earth as [acid rain](#), having a number of [adverse effects](#) on the environment and human life.[59]
- Aerosols interact with the Earth's energy budget in two ways, directly and indirectly.
  E.g., a *direct* effect is that aerosols scatter and absorb incoming solar radiation.[60] This will mainly lead to a cooling of the surface (solar radiation is scattered back to space) but may also contribute to a warming of the surface (caused by the absorption of incoming solar energy).[61] This will be an additional element to the [greenhouse effect](#) and therefore contributing to the global climate change.[62]

  The *indirect* effects refer to the aerosols interfering with formations that interact directly with radiation. For example, they are able to modify the size of the cloud particles in the lower atmosphere, thereby changing the way clouds reflect and absorb light and therefore modifying the Earth's energy budget.[59]
- When aerosols absorb pollutants, it facilitates the deposition of pollutants to the surface of the earth as well as to bodies of water.[62] This has the potential to be damaging to both the environment and human health.
- Aerosol particles with an effective diameter smaller than 10 μm can enter the bronchi, while the ones with an effective diameter smaller than 2.5 μm can enter as far as the gas exchange region in the lungs,[63] which can be hazardous to human health.
-
- **What is the aerosol effect?**
- **Aerosols** and Clouds (Indirect **Effects**) Whereas **aerosols** can influence climate by scattering light and changing Earth's reflectivity, they can also alter the climate via clouds. On a global scale, these **aerosol** "indirect **effects**" typically work in opposition to greenhouse gases and cause cooling.
-
- **What are aerosols give two examples?**
- An **aerosol** is a colloid of fine solid particles or liquid droplets, in air or another gas. **Aerosols** can be natural or artificial. **Examples** of natural **aerosols** are fog, geyser steam. **Examples** of artificial **aerosols** are haze, dust, particulate air pollutants and smoke.
-
- **Where are aerosols found?**
- These ubiquitous specks of matter are known as **aerosols**, and they can be **found** in the air over oceans, deserts, mountains, forests, ice, and every ecosystem in between

What happens when aerosol amounts increase?
An **increased amount** of **aerosols** may **increase** the CCN number concentration and lead to more, but smaller, cloud droplets for fixed liquid water content. This **increases** the albedo of the cloud, resulting in enhanced reflection and a cooling effect, termed the cloud albedo effect

What are the major sources of aerosols?
Natural sources of aerosols include sea salt generated from breaking waves, mineral **dust** blown from the surface by wind, and organic aerosols from biogenic emissions. Artificial, also called anthropogenic, aerosols include sulfate, nitrate, and carbonaceous aerosols, and are mainly from fossil fuel combustion sources.

Aerosols and their Relation to Global Climate and Climate Sensitivity

Atmospheric aerosols from human activity influence climate. Uncertainties in the understanding of their effects limit our knowledge about climate change.

What is the source and composition of the aerosols?

Atmospheric aerosols are suspensions of liquid, solid, or mixed particles with highly variable chemical composition and size distribution (Putaud *et al.* 2010). Their variability is due to the numerous sources and varying formation mechanisms (Figure 1). Aerosol particles are either emitted directly to the atmosphere (primary aerosols) or produced in the atmosphere from precursor gases (secondary aerosols).

Primary aerosols consist of both inorganic and organic components. Inorganic primary aerosols are relatively large (often larger than 1 µm) and originate from sea spray, mineral dust, and volcanoes. These coarse aerosols have short atmospheric lifetimes, typically only a few days. Combustion processes, biomass burning, and plant/microbial materials are sources of carbonaceous aerosols, including both organic carbon (OC) and solid black carbon (BC). BC is the main anthropogenic light-absorbing constituent present in aerosols. Its main sources are the combustion of fossil fuels (such as gasoline, oil, and coal), wood, and other biomass. Primary BC and OC containing aerosols are generally smaller than 1 µm.

Secondary aerosol particles are produced in the atmosphere from precursor gases by condensation of vapours on pre-existing particles or by nucleation of new particles. A considerable fraction of the mass of secondary aerosols is formed through cloud processing (Ervens *et al.* 2011). Secondary aerosols are small; they range in size from a few nanometres up to 1 µm and have lifetimes of days to weeks. Secondary aerosols consist of mixtures of compounds; the main components are sulphate, nitrate, and OC. The main precursor gases are emitted from fossil fuel combustion, but fires and biogenic emissions of volatile organic compounds (VOCs) are also important. Occasionally volcanic eruptions result in huge amounts of primary and secondary aerosols both at the ground and in the stratosphere (Boulon *et al.* 2011).

The size and chemical composition of the particles evolve with time through coagulation, condensation, and chemical reactions. Particles may grow by uptake of water, a process that depends on chemical composition, particle size, and ambient relative humidity. The different particles have varying impacts in the atmosphere depending on composition, and the numerous sources and large range in size distributions further complicate a quantification of their effects. Both particle growth and the mixing of different particle types influence the climate effect of aerosols.

How are aerosols distributed globally?

Aerosol optical depth (AOD) retrieved by remote sensing from space is highly inhomogeneous, with the largest values in Asia and the tropical regions of Africa (see Figure 2). The estimated contributions from different aerosol types in selected regions are shown in pie charts. In general, there is large spatial and temporal variability in global aerosol composition. Remote sensing from both space and ground together with *in situ* observations have substantially advanced an understanding of geographical aerosol distribution, but there are still large uncertainties in the chemical composition and the anthropogenic contribution to the AOD (Figure 2).

Figure 2: MODIS aerosol optical depth.
MODIS aerosol optical depth [AOD (550 nm); dimensionless] averaged over the 10-year period 2001–2010 (Remer *et al*. 2008). Pie charts show how various aerosol types contribute to the total AOD for different regions, as estimated by a global aerosol model (Myhre *et al*. 2009). Aerosol types are Sul (sulphate), BC and OC from fossil fuel usage, Bio (OC and BC from biomass burning), Nitrate, Sea (sea salt), and Min (mineral dust). Gray areas indicate lack of MODIS data. Some aerosol types, e.g. sulphate, have enhanced contributions to AOD due to hygroscopic growth. The contribution from OC is likely underestimated as in most of the global aerosol models (Zhang *et al*. 2007).

How do aerosols affect the climate?

All atmospheric aerosols scatter incoming solar radiation, and a few aerosol types can also absorb solar radiation. BC is the most important of the latter, but mineral dust and some OC components are also sunlight absorbers. Aerosols that mainly scatter solar radiation have a cooling effect, by enhancing the total reflected solar radiation from the Earth. Strongly absorbing aerosols have a warming effect. In the atmosphere, there is a mixture of scattering and absorbing aerosols, and their net effect on Earth's energy budget is dependent on surface and cloud characteristics. Scattering aerosols above a dark surface and absorbing aerosols above a bright surface are most efficient (see Figure 3a). Scattering (absorbing) aerosol above a bright (dark) surface are less efficient because the solar radiation is reflected (absorbed) anyway. Absorbing aerosols are particularly efficient when positioned above clouds, which are a main contributor to the total reflection of solar radiation back to space.

Figure 3: The direct aerosol effect and the cloud albedo effect.
(a) The direct aerosol effect for low and high surface albedo, for scattering and absorbing aerosols. A dark surface (low albedo) will already absorb a large portion of the solar radiation, and absorbing aerosols will thus have a small effect. Scattering aerosols will instead amplify the total reflectance of solar radiation, since the solar radiation would otherwise be absorbed at the surface. Over a bright surface (high albedo) scattering aerosols have a reduced effect. Absorbing aerosols may, however, substantially reduce the outgoing radiation and thus have a warming effect. (b) The cloud albedo effect (first indirect aerosol effect), cloud lifetime effect (second indirect aerosol effect), and semi-direct effect.

Aerosols are vital for cloud formation because a subset of them may serve as cloud condensation nuclei (CCN) and ice nuclei (IN). An increased amount of aerosols may increase the CCN number concentration and lead to more, but smaller, cloud droplets for fixed liquid water content. This increases the albedo of the cloud, resulting in enhanced reflection and a cooling effect, termed the cloud albedo effect (Twomey 1977; Figure 3b). Smaller drops require longer growth times to reach sizes at which they easily fall as precipitation. This effect, called the cloud lifetime effect, may enhance the cloud cover (see illustration in Figure 3b) and thus impose an additional cooling effect (Albrecht 1989). However, the life cycles of clouds are controlled by an intimate interplay between meteorology and aerosol-and-cloud microphysics, including complex feedback processes, and it has proven difficult to identify the traditional lifetime effect put forth by Albrecht (1989) in observational data sets.

Absorbing aerosols also have the potential to modify clouds properties, without directly acting as CCN and IN, by: (1) heating the air surrounding them while reducing the amount of solar radiation reaching the ground, which stabilizes the atmosphere and diminishes the convection and thus the potential for cloud formation, (2) increasing the atmospheric temperature, which reduces the relative humidity, inhibits cloud formation, and enhances evaporation of existing clouds. This is collectively termed the semi-direct aerosol effect (Hansen *et al.* 1997). The net effect is uncertain (see Figure 3b) and highly depends on the vertical profile of BC (Koch & Del Genio 2010).

In addition, BC and other absorbing aerosols deposited on snow or ice surfaces may reduce the surface albedo, leading to reduced reflectance of solar radiation, and hence a heating effect (Hansen & Nazarenko 2004).

Radiative forcing (RF) is often used to quantify and compare the potential climate impact of the various aerosol effects. RF is defined as a change in the Earth's radiation balance due to a perturbation of anthropogenic or natural origin.. The total aerosol forcing probability density function (PDF), in addition to individual aerosol components, indicating both the magnitudes and uncertainty of the effects, is shown in Figure 4a. The wider a PDF, the larger is the uncertainty. Combining all aerosol effects (blue dashed curve in Figure 4a) enhances the uncertainty compared to considering only the direct aerosol effect and cloud albedo effect.

(a) Probability density functions of aerosol effects (Isaksen *et al.* (2009), with small updates of cloud albedo and lifetime effects). The total aerosol radiative forcing (red and blue curves), with and without clouds are estimated by combining the individual effects in a Monte Carlo calculation (Boucher & Haywood 2001). Vertical lines show 90% confidence intervals. (b) Climate sensitivity for a doubling of CO_2 as a function of the total aerosol RF. Radiative imbalances of 0.85 (solid line, Hansen *et al.* 2005), 0.7 and 1.0 Wm^{-2} (grey band) and 0.0 (radiative equilibrium, dashed line) are shown. Industrial era temperature change is taken as 0.8 Kelvin (K), and RF of non-aerosol components +2.9 Wm^{-2}.

Why is the uncertainty in the aerosols important for predictions of climate sensitivity?

The change in global mean surface temperature (ΔT) as a result of RF can be expressed by the following simple heat-balance equation:

$$c \quad d(\Delta T)/dt \quad = \quad RF \quad - \quad \Delta T/\lambda \qquad (1)$$

Here *c* is the heat capacity of the land-ocean-atmosphere system and λ is the climate sensitivity. At radiative equilibrium (d(ΔT)/dt = 0), Equation 1 reduces to $\Delta T = \lambda RF$. However, the Earth is not in radiative equilibrium, since less thermal radiation is currently emitted to space compared to what is absorbed of solar radiation (Hansen *et al.* 2005). This radiative imbalance causes the Earth to gradually warm, with global warming as a result (Trenberth & Fasullo 2010). The simple equation above has two key uncertainties. The observed surface temperature change is rather well determined over the industrial era, but the climate sensitivity and the total RF are both highly uncertain. The climate sensitivity is an essential parameter for prediction of future climate change. Quantifying the climate sensitivity for the doubling of CO_2 has long been attempted by using global climate models or temperature records, but it still has a wide range of reported values (IPCC 2007, Knutti & Hegerl 2008). The total RF through the industrial era is also uncertain, mainly due to lack of quantification of the aerosol effects discussed above. The implication of this uncertainty in the aerosol RF for the quantification of the climate sensitivity can be illustrated as follows:

If we assume a total aerosol RF and a current energy imbalance, we can compute the resulting climate sensitivity using Equation 1 (Figure 4b). This can then be compared with the PDFs for the current aerosol RF to get an indication of the range in climate sensitivities allowed by the present knowledge (red and blue lines in figure 4b). A similar figure has previously been presented in Andreae *et al.* (2005). The allowed climate sensitivity ranges from about 2 to 8 Kelvin (K) for a doubling of CO_2 using the known industrial age warming of around 0.8 K, the present best knowledge of RF from non-aerosol components, the 90% confidence interval of the total aerosol RF for the most certain effects, and radiative imbalance.

Has there been any progress in the understanding of the climate effect of aerosols?

There has been a tremendous improvement in the understanding of atmospheric aerosols and their climate effect over the last decades, with some important observational and modelling breakthroughs. Long-term measurements of aerosols (e.g., Putaud *et al.* 2010, Andrews *et al.* 2011), observational campaigns (e.g., Quinn & Bates 2005), and remote sensing from space and ground (Holben *et al.* 1998, Remer *et al.* 2008) have remarkably increased knowledge about the composition and characteristics of atmospheric aerosols. However, an understanding of the greater complexity of atmospheric aerosols has at the same time limited more robust quantification of their climate effect. The first estimate of the direct aerosol effect in the early 1990s was limited to sulphate aerosols (Charlson *et al.* 1991), with estimates for BC coming a few years later (Haywood & Shine 1995). Observations have shown that OC is an important aerosol component (Novakov *et al.* 1997, Ramanathan *et al.* 2001), and substantial investigations have later explored the complex composition and optical characteristics of this compound (e.g., Kanakidou *et al.* 2005, Graber & Rudich 2006). Global aerosol models today provide RF estimates for a large set of aerosol components, such

as sulphate, BC (from fossil fuel and biomass burning), OC (primary and secondary from fossil fuel and biomass burning), and nitrate (Jacobson 2001, Liao & Seinfeld 2005, Koch *et al.* 2009). In addition, multi-model studies are performed to understand and reduce uncertainties due to model differences (Schulz *et al.* 2006).

An example of recent progress is reduced uncertainty in the estimate of the total direct aerosol effect. This estimate was made possible by advances that have occurred on both the modelling and the observational side, and was based on a combination of global aerosol models and observation based methods (mostly remotely sensed data). Initially, observational estimates of RF were up to three times stronger than model based calculations (Forster *et al.* 2007). Consistency between these two different approaches has subsequently been reached, and was found to arise from necessary and simplified assumptions of the pre-industrial aerosol composition in the observation-based method (Myhre 2009). Although the uncertainty in the total direct aerosol effect is reduced, it is still substantial compared to uncertainties associated with greenhouse gases. In addition the uncertainty in individual RF for several of the aerosol components, such as BC, OC, and nitrate, is large.

Similar to the early estimates of the direct aerosol effect, many of the first model estimates of the aerosol indirect effect only accounted for the effect of sulphate particles acting as CCN (Kaufman & Chou 1993, Jones *et al.* 1994). Furthermore, they only included the influence of sulphate aerosols on cloud albedo, disregarding any effects on cloud lifetime and extent. With the realization that other aerosol species of anthropogenic origin could also form cloud droplets and that effects on cloud lifetime and extent were also possible, global climate models estimated the aerosol indirect effect to be stronger (e.g., Lohmann & Feichter 1997, Menon *et al.* 2002). Some even predicted this cooling effect to be comparable in magnitude to the warming greenhouse effect. Recent publications have later pointed to oversimplifications in model representation of clouds and how their lifetimes are affected by aerosols (e.g., Stevens & Feingold 2009). It is now acknowledged that aerosol effects on cloud lifetime will vary with the cloud type in question, and that complex feedback processes can sometimes complicate the ultimate cloud response to aerosol perturbations. Recent model studies have found that by forming ice in super-cooled liquid clouds, aerosols may in fact shorten cloud lifetime, because of the more efficient precipitation formation when cloud ice is present (e.g., Lohmann & Hoose 2009, Storelvmo *et al.* 2011). In summary, whether aerosols are acting as CCN or IN or are simply modifying atmospheric stability by absorbing solar radiation, there is still high uncertainty associated with their effect on cloud lifetime. This uncertainty reflects how challenging it is to represent aerosol-and-cloud processes that occur on microscopic scales in models that have resolutions of tens to hundreds of kilometres. Although much uncertainty remains, model and satellite estimates of the cloud albedo effect seem to converge on a negative RF that has about half the magnitude of the positive RF attributed to increasing CO_2 concentrations.

What are the different types of aerosols?
Primary **aerosols** contain particles introduced directly into the gas; secondary **aerosols** form through gas-to-particle conversion. **Various types of aerosol**, classified according to physical form and how they were generated, include dust, fume, mist, smoke and fog. There are several measures of **aerosol** concentration.

Aerosols and climate

What are aerosols?

Aerosols are fine, airborne particles consisting at least in part of solid material. Density of the basic materials of aerosols range from 1.0 g/cm^3 (for soot) to 2.6 (for minerals). The ocean is a major source of natural aerosols. Air-sea exchange of particulate matter contributes to the global cycles of carbon, nitrogen, and sulfur aerosols, such as dimethylsulfide (DMS) produced by phytoplankton. Ocean water and sea salt are transferred to the atmosphere through air bubbles at the sea surface. As this water evaporates, the salt is left suspended in the atmosphere. Four other significant sources of aerosols are terrestrial biomass burning,

volcanic eruptions, windblown dust from arid and semi-arid regions, and pollution from industrial emissions (**Fig 1**).

Clean continental air often contains less than 3,000 particles per cubic centimetre (of which half are water-soluble), polluted continental air typically 50,000/cm^3 (of which two-thirds are soot, and the rest mostly water-soluble). Urban air typically contains 160,000/cm^3, mostly soot, and only 20% is water-soluble. Desert air has about 2,300/cm^3 on average, almost all water-soluble. Clean marine air generally has about 1,500/cm^3, about all water-soluble. The lowest sea-level values occur over the oceans near the subtropical highs (600/cm^3 on average, but occasionally below 300/cm^3). Arctic air has about 6600/cm^3 (including 5,300 soot) and on the Antarctic plateau only 43/cm^3 occur (about all sulphate) (1).

Sources of the key climate-affecting aerosol types

Source	Region	Dominant Species
Industrial pollution	Eastern North America, Europe, Eastern Asia	water-soluble inorganic (e.g., sulfate, nitrate, ammonium), organic carbon, elemental carbon
Biomass combustion	tropical/subtropical South America and Africa	organic carbon, elemental carbon
Wind-blown dust	disturbed arid soils	mineral dust
Natural	remote continental, remote marine, free troposphere	

Fig 1. Main sources and types of aerosols that affect climate.

Concentrations of aerosols decrease about exponentially with height, thus

$$N(z) = N(0) \exp(-z/H)$$

where N(z) is the concentration at height z (km) above ground level. The scale height H typically is 10 km for continental (incl desert) air in summer, 6 km for urban air, 2 km for continental air in winter, and 1 km over the marine subtropical high pressure regions. The higher values during the summer imply more deep-tropospheric stirring of the aerosols by thunderstorms. In places with a well-defined, long-lived inversion on top of a well-mixed planetary boundary layer, the aerosol concentration tends to be homogenous in the PBL and perhaps 10x lower above the inversion.

Aerosols can be long-lived and therefore be advected over a long distance. Four weeks of measurements of tropospheric aerosols at Mildura (in the Australian state of Victoria) showed that some came from Africa, about eight days upwind, and possibly South America, even further away. These aerosols may have come from gases generated by forest fires (2).

How aerosols affect climate

Aerosols play an important role in the global climate balance, and therefore they could be important in climate change. Natural variations of aerosols, especially due to episodic large eruptions of volcanoes, such as Mt. Pinatubo in 1991, are recognized as a significant *climate forcing*, that is, a factor that alters the Earth's radiation balance and thus tends to cause a global temperature change. In addition, there are several ways in which humans are altering atmospheric aerosols, not only near the ground (e.g. industrial emissions) but as

high as the lower stratosphere (where they are continuously emitted by aircraft), and thus possibly affecting climate (e.g. through contrails) (3).

Aerosols force climate in two ways (4):

- **direct** radiative forcing: the scattering of solar radiation and the absorption/emission of terrestrial radiation.
- **indirect** radiative forcing, mainly by effects of aerosols on cloud properties. (A minor indirect effect involves the heterogeneous chemistry of greenhouse gases: these gases may react at the surface of an aerosol and therefore change radiative properties.)

Greenhouse gases have a well-understood effect on the global radiative balance and surface temperatures, their concentration has little variability (except water vapour and ozone), and their long-term trends are well-known. Therefore, there is much confidence in the greenhouse gas component of the anticipated climate change during the next few decades. However climate forcings due to aerosols are not determined well, especially the indirect radiative forcing. Indeed, aerosols are one of the greatest sources of uncertainty in interpretation of climate change of the past century and in projection of future climate change.

The effect of aerosols on clouds is highly speculative. The theory is that the more aerosol, the smaller the cloud droplets tend to be, and clouds with more but smaller drops have a higher albedo. This would increase the planetary albedo, i.e. have a cooling effect. Twomey (5) proposed the first step in this theory, by showing empirically that the mean droplet mass in a cloud decreases in proportion with the number concentration of aerosols (N). In other words, the mean droplet radius (r) is proportional to:

$$r \sim N^{-0.33} \text{ or } \Delta r/r = -(\Delta N/N)/3$$

So if the aerosol concentration increases by 30% ($\Delta N/N = 0.3$), the droplet radius decreases by 10%, and for the same total cloud water content, the number of droplets will increase by 30%.

While this relation has been corroborated surprisingly well in various field experiments, it is uncertain whether the cloud water content will be conserved. And the radiative forcing of clouds depends strongly on the heights of their bases and tops, which are unknown. Also, a variation in the drop size of a cloud will affect the way in which the cloud will evolve, produce rain or evaporate. In some circumstances aerosols may create clouds where none existed before, because they act as cloud condensation nuclei. Satellite imagery over ocean areas prone to stratus clouds (especially over high-pressure regions and/or low sea surface temperatures) reveal that ships often trigger lines of stratus along their track. These 'ship tracks' increase the albedo and cause net cooling. Another example is the production of DMS by phytoplankton (3). Some of this DMS seeps into the atmosphere where the sulphate aerosol enhances cloud formation. Accurate measurements at Cape Grim in northwest Tasmania have shown that DMS is the main agent of the nucleation of clouds over the southern oceans (6). The surface waters within the large ocean gyres are generally depleted of nutrients, especially iron. An experiment has shown that the seeding of iron dramatically increases the DMS production in the ocean. The fractions of aerosols that are manmade, biogenic, volcanic, or soil-based, and the chemical reactions of some aerosols, are poorly understood. The inclusion of the ice phase adds an extra step of complexity. In short, the effect of aerosols on clouds and thus on climate is very uncertain.

 The climate forcing of greenhouse gases, aerosols, and other variants, is commonly expressed in terms of the resulting net change to the radiation balance, expressed in W m^{-2}. What really matters of course is how much warming or cooling a variant bears. For instance, the net warming of the Earth's atmosphere (as compared to an atmosphere without greenhouse gases, aerosols or clouds) is 33K at the surface (from -18 C to +15 C, Section 2.8). The reason for the use of radiation units rather than temperature units is that the effect of some perturbation (such as a volcanic eruption) on surface temperatures involves a cornucopia of complex feedback effects, such as atmospheric and ocean circulations. The radiative forcing is more 'raw', and allows more direct comparisons, e.g. between general circulation models (GCMs).

The aerosol concentrations depend on the wind, the land surface conditions (vegetation cover), sea surface temperature, and other climate factors. During an Ice Age the aerosol types and distributions were very different from those during an interglacial. It is possible that aerosol acts as a positive feedback in enhancing the difference between glacials and interglacials. For instance, during a glacial, the stronger winds over denuded periglacial plains might pick up more dust, contributinging to a higher planetary albedo, maintaining the cold. In other words, climate and aerosol are inter-related in a complex way.

Measuring aerosol concentrations

There are many surface stations that measure at least some component of aerosols (their size, type, concentration). These measurements tend to be biased by the local/regional industrial activities, vehicle transport, land surface cover and soil type, and winds. There are only few stations that are truly remote from sources of manmade aerosols, e.g. the Cape Grim Baseline Air Pollution Station in Tasmania, or the Mauna Loa Observatory in Hawaii. In short, in view of local sources, the high spatial/temporal variability of aerosols, and the complexity of aerosol properties, aerosols are much undersampled to really understand their distribution.

Aerosol could be sampled remotely, from satellite, but only indirectly, by means of a combination of various wavelengths of upwelling radiation (7). Even then numerous assumptions are required. At this time there is no satellite instrument that is specifically designed to measure aerosol. There are two instruments that can be used to estimate the spatial distribution of aerosol optical thickness, but not their vertical profile. The optical thickness is an integrated amount, and in the visible spectrum it is a measure of haziness.

One satellite instrument is the AVHRR (Advanced Very High Resolution Radiometer). A combination of spectral bands was used to estimate the global-mean aerosol *optical thickness,* shown in Fig 2 (8). An optical thickness of 1 means that no radiation (in this case light at 0.55 micron) directly penetrates through the atmosphere. The high values at low latitudes (esp. around Africa) may be largely due to fires in support of subsistence farming. The AVHRR-based technique performs poorly over land because of the high variability of radiative properties of the land surface. Even over the ocean the signal-to-noise ratio is not much more than one. The other instrument is TOMS (Total Ozone Mapping Spectrometer). TOMS data can be used to detect absorbing aerosols over land, but are insensitive to aerosols located below 1 km, including the planetary boundary layer in which most aerosols reside.

More accurate aerosol measurements will soon be taken by the MODIS instrument to be launched soon on the Terra satellite. MODIS (Moderate Resolution Imaging Spectrometer) will measure global variations in aerosol optical thickness over land and oceans, as well as global aerosol particle size distribution over oceans (as there is currently no simple way to derive aerosol particle size distribution over land). The algorithms for remote sensing of aerosols over oceans and land (MODIS Aerosol Product) are very different from one another because of differences in the spectral reflectance of water and land under the semi-transparent aerosol layer. MODIS's objective is to provide a comprehensive series of high-resolution global observations of the Earth's land, oceans, and atmosphere in the visible and infrared regions of the spectrum in such a way as to view the entire surface of the Earth every two days. MODIS will continue to take measurements in spectral regions that have been and are currently being measured by other satellite sensors, such as the AVHRR (9).

Recently (1998) the Global Aerosol Climatology Project was established to obtain a better insight of aerosol concentrations and types, by means of satellite and surface data, field experiments and numerical simulations. The movement and chemical evolution of aerosols can be simulated by means of GCMs that include cloud microphysics, but until recently these models have used crude assumptions, and they could be initialized better by means of satellite-retrieved data.