Chapter 11

100 Years of Progress in Cloud Physics, Aerosols, and Aerosol Chemistry Research

SONIA M. KREIDENWEIS
Department of Atmospheric Sciences, Colorado State University, Fort Collins, Colorado

MARKUS PETTERS
Department of Marine Earth and Atmospheric Sciences, North Carolina State University, Raleigh, North Carolina

ULRIKE LOHMANN
Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

ABSTRACT

This chapter reviews the history of the discovery of cloud nuclei and their impacts on cloud microphysics and the climate system. Pioneers including John Aitken, Sir John Mason, Hilding Köhler, Christian Junge, Sean Twomey, and Kenneth Whitby laid the foundations of the field. Through their contributions and those of many others, rapid progress has been made in the last 100 years in understanding the sources, evolution, and composition of the atmospheric aerosol, the interactions of particles with atmospheric water vapor, and cloud microphysical processes. Major breakthroughs in measurement capabilities and in theoretical understanding have elucidated the characteristics of cloud condensation nuclei and ice nucleating particles and the role these play in shaping cloud microphysical properties and the formation of precipitation. Despite these advances, not all their impacts on cloud formation and evolution have been resolved. The resulting radiative forcing on the climate system due to aerosol–cloud interactions remains an unacceptably large uncertainty in future climate projections. Process-level understanding of aerosol–cloud interactions remains insufficient to support technological mitigation strategies such as intentional weather modification or geoengineering to accelerating Earth-system-wide changes in temperature and weather patterns.

1. Introduction: Aerosols, clouds, and precipitation

Suspended particles are ubiquitous in Earth’s atmosphere and have long been recognized to affect visibility and Earth’s radiative balance. Their number, size, and chemical composition vary widely with meteorology, season, elevation, and geographic location. Particles suspended in the atmosphere range in size between 1 nm and 10 μm; larger particles can also become airborne but have relatively short lifetimes due to their high gravitational settling velocities. Atmospheric particles have been shown to have constituents that include sulfate, nitrate, chloride, and other salts, minerals, biological material, as well as graphitic and organic carbon, from both primary sources (directly emitted) and secondary sources (formed in the atmosphere via chemical and physical processes). Individual particles are often morphologically complex mixtures of several different components that make up the aerosol (e.g., Takahama et al. 2007, 2010; Bzdek et al. 2012).

The interactions of aerosols in the troposphere and stratosphere with incoming solar radiation had been pointed out early in the twentieth century. Charlson...
et al. (1992), building on the estimates of Charlson et al. (1990), argued that the addition of anthropogenic aerosols to the climate system constituted a significant perturbation—a climate forcing—to the natural system. Their analysis focused on sulfate aerosol, due to anthropogenic emissions that could be shown to significantly exceed natural sources by ~1950 (Fig. 11-1). Their simple but elegant numerical analysis suggested that “anthropogenic sulfate produces a mean radiative forcing of the NH [Northern Hemisphere] climate that is comparable in magnitude but opposite in sign to the anthropogenic perturbation in forcing by greenhouse gases,” the latter phenomenon having been a focus of the scientific community since the 1950s and the publication of Keeling’s data in the 1960s (Wallington et al. 2019). In an early global modeling study, Kiehl and Briegleb (1993) demonstrated the spatial variability in greenhouse gas and aerosol forcing (Fig. 11-2), showing that the “offset” of greenhouse gas warming by aerosol cooling was not a simple cancellation of effects, and would have impacts on atmospheric dynamics due to changes in gradients of heating as also foreseen by Charlson et al. (1990).

Charlson et al. (1992) included in their analysis of aerosol climate forcing a discussion of the “indirect radiative influence” of sulfate aerosols that arises due to the interactions of atmospheric particles with cloud formation and development. Sufficiently large numbers of particles are always present in the atmosphere, such that all water vapor condensation is facilitated by particles. The formation of cloud droplets and ice crystals (at $T > -38^\circ$C) is therefore strongly tied to the nature of the atmospheric aerosol. Research by Aitken, Köhler, Bergeron, Findeisen, and Schaefer in the 1910–40s revealed that cloud condensation nuclei (CCN) and ice nucleating particles (INP) were specific subsets of the total atmospheric particulate matter. Their potential importance for cloud physics was understood immediately.

Figure 11-3 shows a recent time history of the number of papers on CCN and INP. Starting in the late 1950s there was significant interest in ice nucleation research, in part due to the hope of successful weather modification by...
cloud seeding. At the same time there was less interest in the nature of CCN. Twomey (1977) was a landmark article linking pollution, CCN, and planetary albedo, which is now referred to as the Twomey effect, an indirect aerosol effect (or radiative forcing) on climate due to aerosol–cloud interactions (Boucher et al. 2013). Twomey’s work showed that all else being equal, an increase in CCN number leads to more numerous and smaller cloud droplets, and in turn more reflective clouds. Ten years later, Charlson, Lovelock, Andreae, and Warren (CLAW; Charlson et al. 1987) proposed a biogeochemical climate feedback loop involving CCN (Fig. 11-4), which was subsequently named the CLAW hypothesis. The hypothesis is that dimethyl sulfide (DMS) produced by phytoplankton results in the formation of CCN and an increase of cloud albedo through the Twomey effect. The increased cloud albedo then cools the planet by blocking incoming solar radiation. Twomey and CLAW gave birth to the field of study of aerosol–cloud interactions. Starting in the early 1990s, a sharp increase in the number of publications on the topic occurred (Fig. 11-3), followed by an approximately exponential growth that continues to this day.

The CCN spectrum is amenable to direct measurement using CCN counters (Twomey 1959b; Hudson and Squires 1976; Nenes et al. 2001a). Figure 11-5 illustrates early measurements of the resulting cumulative CCN activation spectra that consist of CCN number as a function of supersaturation. No size or chemical composition information is required. The spectra can then be used directly to predict the number of cloud droplets in adiabatic updrafts via simple parameterizations (Cohard et al. 1998; Twomey 1959b). Simplifications include the lack of explicit treatment of aerosol water uptake, kinetic limitations to CCN activation, variations in updraft velocity, and modification of the CCN by the atmospheric trace gas pool. Nonetheless, the concept of the CCN spectrum serves as an intermediate step that divides the aerosol–cloud interaction problem into two steps: the link between aerosol size distribution/chemical composition and CCN number and the link between CCN number and cloud droplet number concentration.

Similar arguments can be made about ice formation. For clouds warmer than −38°C, INPs are needed to initiate ice below and above water saturation. INPs are generally water insoluble or crystalline at the prevailing conditions, and observations suggest that particles having diameters larger than 500 nm dominate the sources (DeMott et al. 2010). Nucleation occurs at some supercooling at the surface of the INP, and the process has a stochastic component introducing a weak but measurable time dependence into the process. In contrast to the CCN problem, first principle theories that link particle chemical composition or particle surface properties and ice nucleation are not well established. Parameterizations based on laboratory and field data are therefore used to compute INP concentrations from aerosol composition data.

Analogous to the CCN activation spectrum, cumulative INP spectra as functions of supercooling can be

---

**Fig. 11-3.** Number of publications returned by Clarivate Web of Science for the following searches: “topic” containing aerosol and climate; the title or topic containing “cloud condensation nuclei” or “CCN”; and the title or topic containing “ice nucleation” or “ice nuclei.”

**Fig. 11-4.** Illustration of the proposed CLAW climate feedback loop. Rectangles are measurable quantities and circles are processing links. [Figure from Charlson et al. (1987). Reprinted by permission from Springer Nature.]
constructed from experimental data. Figure 11-6 shows composite INP spectra for immersion mode freezing nucleation that illustrate the wide range of INP number concentrations that are available in the atmosphere. At $-20^\circ$C, INP number concentrations are typically five to six orders of magnitude smaller than CCN concentrations. As with the CCN spectra, INP spectra are amenable to direct measurement and can be used to predict a lower bound of ice crystals that is expected at a given temperature in a cloud. Actual ice crystal concentrations inside the cloud may be much larger due to secondary ice processes (see Field et al. 2017). Simplifications of inferring first ice formation from immersion mode INP spectra include the inability to represent the time dependence of the nucleation process, the dynamic preferential removal of INPs due to cold-rain formation, and the inclusion of a wide range of freezing processes that are not represented in this framework. Again, analogous to the CCN spectrum, the INP spectrum serves as an intermediate step that divides the aerosol–ice cloud interaction problem into the link between aerosol particle size, particle surface properties, and chemical composition and the INP spectrum, and the link between the INP spectrum and ice cloud properties.

For both liquid and ice clouds, the initial formation of cloud particles is then followed by processes that include interactions with water vapor, interactions between cloud particles, and interactions between cloud and precipitation particles, all shaping the microphysical evolution and formation of precipitation. The recognition and development of cloud physics as a distinct field of study, and its relevance to the field of numerical weather prediction, is credited to Sir John Mason (Browning 2016). Understanding of cloud microphysical processes was first developed through laboratory efforts. Although laboratory-based research continues to the present (Chang et al. 2016), the advent of airborne platforms and associated novel measurement techniques—both in situ and via remote sensing—accelerated the viability of field studies that could probe more realistic conditions for the broad variety of cloud types, ranging from fog and stratus to deep tropical convection. It was also realized early on that dynamics and microphysics were strongly coupled, and that modeling efforts must represent both processes. Early developments of coupled models were strongly motivated by interest in weather modification (Cotton and Pielke 1976), essentially attempts to influence cloud microphysical processes to affect precipitation; more recently, inadvertent
weather modification due to human activities that perturb these processes has been a focus of global change research. Representations of these microphysical processes in models, for scales ranging from single parcels, to clouds and cloud systems, and ultimately at global scales, thus seek to accurately capture factors that link aerosols, clouds, precipitation, cloud and cloud environment dynamical processes, and the hydrological cycle. Figure 11-7 depicts the overview by Hobbs (1991) of the major cloud physics research focus areas by decade, beginning in the 1950s, at which point the basic theories of droplet and ice crystal nucleation, condensation, collection, and precipitation initiation had already been developed. New developments in recent decades have been added to Fig. 11-7 by the authors.

In the following sections, we describe the evolution over the last century in scientific understanding of the atmospheric aerosol, aerosol and cloud microphysics, and aerosol–cloud interactions. The latter topic includes ongoing efforts in understanding and quantifying anthropogenic effects on climate that are induced by aerosols and by aerosol–cloud–precipitation interactions.

2. The atmospheric aerosol

Volcanic eruptions, dust storms, and smoke are highly visible manifestations of suspended particulate matter, but the presence of smaller particles in the atmosphere and their importance to cloud formation was deduced as early as 1875, when Coulier reported the results of expansion chamber experiments. His studies showed that unfiltered air condensed water droplets more readily than filtered air, clearly demonstrating “the important part played by dust in the cloudy condensation of the vapor in air” (Aitken 1881). Aitken was conducting parallel experiments at the time, also deducing the importance of “dusts” in cloud formation. He further asserted that the atmosphere rarely contained fewer than about 100 particles per cubic centimeter (Aitken 1909)—that is, that all cloud droplet formation occurred on particulate nuclei, as they were ubiquitous. At the time that Simpson (1941) published his article “On the formation of cloud and rain,” it was understood that both hygroscopic and nonhygroscopic “condensation nuclei” (defined by the measurement methodologies; see below) existed in the atmosphere, with three types proposed: products of combustion, or “smoke,” that included condensed sulfuric and nitrous acids; insoluble particles, including mineral dust and insoluble combustion products; and various salts originating from sea spray. The work of Christian Junge (1912–96; e.g., Junge 1953) and others further clarified the main sources, sizes, and composition that composed the atmospheric aerosol.

Remarkable advancements in understanding of the sources of atmospheric particulate matter and the physical and chemical properties of particles have occurred over the last century, in large part due to the development of new measurement technologies and, importantly, of accepted calibration standards. We cannot possibly do justice to the vast body of knowledge contributed to these topics by researchers worldwide. Here, we present selected historical context and highlights and refer to review articles that present more thorough treatments. In addition to those referenced below, Petzold et al. (2013; chapter 4 in Wendisch and Brenguier 2013), Andreae et al. (2009), and Gong and Barrie (2009; chapters 3 and 4 in Levin and Cotton 2009) provide a recent overview of many in situ aerosol measurement technologies and, in particular, those that have been deployed on airborne platforms to provide descriptions of the spatial distribution of the atmospheric aerosol.

a. Total particle number concentrations

One of the most basic properties of the atmospheric aerosol to be quantified was the number concentration of particles. The first widely used device for this purpose was Aitken’s “dust counter” (Aitken 1890). Detected particles were counted (using a magnifying glass) after forming drops in an expansion-type device. The high supersaturations achieved meant that nearly all particles, down to very small sizes (<10 nm), would be detected, and thus particles detected via such instruments were identified as “condensation nuclei.” Simpson (1941) summarized decades of reported observations of condensation nuclei (Fig. 11-8). The wide range in particle number concentrations clearly demonstrated that urban areas had much higher concentrations than rural and ocean locations. Instruments for counting numbers of particles remain important components of atmospheric research to the present day. McMurry (2000a) presented an historical overview of the development of condensation nuclei counters [also called condensation particle counters (CPCs) or Aitken nuclei counters (ANCs)], along with discussion of their operating principles. These instruments range from the original, water-based expansion devices that were in common use into the 1970s to modern, automated, continuous-flow devices that use various working fluids to grow particles to sizes at which they are readily detected optically by exploiting their light scattering characteristics. The choice of working fluid and the adjustment of the maximum supersaturation enable some discrimination by particle size and type (e.g., Kulmala et al. 2007; Stolzenburg and McMurry 1991).

Condensation particle counters continue to be widely applied in monitoring networks and field campaigns.
### FIG. 11-7. Some decadal milestones since 1950 in research on cloud and precipitation processes. Figure adapted from Hobbs (1991). Milestones for 1990s, 2000s, and 2010s have been added by the authors.

<table>
<thead>
<tr>
<th>Decade</th>
<th>Milestones</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950s</td>
<td>First weather satellite provides global view of clouds</td>
</tr>
<tr>
<td></td>
<td>Understanding of tornadic thunderstorm dynamics through numerical modeling</td>
</tr>
<tr>
<td></td>
<td>Growing interest in aerosol direct and indirect effects</td>
</tr>
<tr>
<td></td>
<td>Convective invigoration by aerosols proposed</td>
</tr>
<tr>
<td></td>
<td>Improved cloud physics probes that eliminate shuttering artifacts</td>
</tr>
<tr>
<td>1960s</td>
<td>Theories of cumulus convection</td>
</tr>
<tr>
<td></td>
<td>Measurements of raindrop size distributions</td>
</tr>
<tr>
<td></td>
<td>Laboratory studies of homogeneous nucleation of ice</td>
</tr>
<tr>
<td></td>
<td>Quantitative treatment of growth of cloud droplets by condensation</td>
</tr>
<tr>
<td></td>
<td>Field and theoretical studies of hailstorms and hail growth</td>
</tr>
<tr>
<td></td>
<td>Measurements of microstructures of clouds</td>
</tr>
<tr>
<td></td>
<td>Early calculations on the growth of raindrops in clouds</td>
</tr>
<tr>
<td></td>
<td>Early radar studies of clouds and precipitation</td>
</tr>
<tr>
<td></td>
<td>Early laboratory and field studies of cloud condensation nuclei and natural and artificial ice nuclei</td>
</tr>
<tr>
<td></td>
<td>Studies of factors affecting shapes of ice crystals</td>
</tr>
<tr>
<td></td>
<td>Early experiments on artificial stimulation of precipitation by cloud seeding</td>
</tr>
<tr>
<td>1970s</td>
<td>Experimental and theoretical studies of collision and breakup of droplets and warm-rain processes</td>
</tr>
<tr>
<td></td>
<td>Laboratory and theoretical studies of Ostwald ripening of soot</td>
</tr>
<tr>
<td></td>
<td>Development of methods for deriving pressure and buoyancy fields in clouds from Doppler radar measurements</td>
</tr>
<tr>
<td></td>
<td>Identification of the utility of color-display Doppler radars for the real-time detection of mesoscale features and severe weather</td>
</tr>
<tr>
<td></td>
<td>Observational and theoretical studies of entrainment in convective clouds</td>
</tr>
<tr>
<td></td>
<td>Reassessments of the potential for increasing precipitation by cloud seeding</td>
</tr>
<tr>
<td>1980s</td>
<td>Development of 3-D numerical models of clouds</td>
</tr>
<tr>
<td></td>
<td>Airborne measurements of the structure and airflow fields in hurricanes</td>
</tr>
<tr>
<td></td>
<td>Recognition of MCC's as a distinct weather system and associated dynamics and kinematics</td>
</tr>
<tr>
<td></td>
<td>Discovery and investigations of &quot;microbursts&quot;</td>
</tr>
<tr>
<td></td>
<td>First approaches using a parameterization of Köhler theory in global climate models</td>
</tr>
<tr>
<td></td>
<td>First estimates of the direct and indirect aerosol effects on warm clouds from global climate models</td>
</tr>
<tr>
<td></td>
<td>Development of nested-resolution 3-D numerical models that included representation of cloud microphysics</td>
</tr>
<tr>
<td></td>
<td>First attempts at simulating prognostic rain and snow in numerical models</td>
</tr>
<tr>
<td></td>
<td>Use of two-moment stratiform cloud microphysics schemes in global climate models</td>
</tr>
<tr>
<td></td>
<td>Resurgence of interest in cloud condensation nuclei</td>
</tr>
<tr>
<td>1990s</td>
<td>First modeling studies of geoengineering with stratospheric aerosols</td>
</tr>
<tr>
<td></td>
<td>First cloud radar in space (CloudSat) and first constellation of satellites aimed at coordinated measurements of cloud properties and processes (A-Train)</td>
</tr>
<tr>
<td></td>
<td>Resurgence of interest in ice nucleation research</td>
</tr>
<tr>
<td></td>
<td>Studies of geoengineering approaches involving clouds and aerosols (marine cloud brightening, cirrus thinning)</td>
</tr>
<tr>
<td></td>
<td>Renewed interest in observations and modeling of microphysical processes and microphysics-dynamics coupling</td>
</tr>
</tbody>
</table>
Modern data confirm Simpson’s summary of the variability and concentration ranges of ambient total particle number concentrations, with observed mean concentrations in the marine boundary layer of 300–2000 cm\(^{-3}\), similar to free tropospheric (mountain sites) observations of 500–2000 cm\(^{-3}\), and mean observations of 800–7000 cm\(^{-3}\) in the continental (nonurban) boundary layer (Spracklen et al. 2010; Fig. 11-9). As shown in Fig. 11-9, number concentration data often show pronounced seasonal cycles that reflect both local/regional sources of particles and variations in meteorology that transport particles from different source regions to the site and that introduce variability in the types and rates of particle deposition processes (e.g., Pierce et al. 2015). Such measurements of particle number concentrations have proven useful in helping elucidate the processes that control new
variable particle formation and growth in the atmosphere, and in particular, in developing understanding of how and where in situ formation of new particles from the gas phase [new particle formation (NPF)] occurs (e.g., Spracklen et al. 2010; Westervelt et al. 2013; Yu et al. 2010).

b. Particle size measurements

Although by the 1970s many observations of atmospheric condensation nuclei existed, it was not until the development of methods to produce precise calibration particles that the scientific value of observational data was fully realized (McMurry 2000b). One of the most important developments in this regard was the use of electrical mobility to size submicron particles (Flagan 1998). Hewitt (1957) published the first description of how charged particles could be sized in an electric field via their electrical mobilities, and this concept was subsequently applied to the development of the differential mobility analyzer (DMA; Langer et al. 1964; Knutson and Whitby 1975; Wang and Flagan 1990; Russell et al. 1995, 1996; Zhang et al. 1995; D.-R. Chen et al. 1998; Liu and Pui 1974; Park et al. 2008; Wang et al. 2017). Differential mobility analyzers are used to produce nearly monodisperse calibration aerosols and thus are used to quantify the detection limits of CPCs. In addition, by using a CPC or other counter as a detection device and scanning through the range of electrical mobilities, the ambient aerosol size distribution can be measured. The upper limit of this technology is generally around 1 μm. DMAs have been designed to measure aerosol size distributions down to ~2 nm, enabling detection of recently formed ultrafine particles as well as the evolution of the submicron number size distribution as particles undergo condensational growth and coagulation.

An early method for sizing individual particles was the use of scattered light intensities, which vary according to particle size; an overview of the development of optical particle counters (OPCs) is provided by Gebhart (1993). Illumination can be by white light or by lasers at various wavelengths. The availability of common light sources restricts the lower limit of detection to particle diameters ~300 nm, although high-intensity light sources and shorter wavelengths can push this limit to lower sizes. A complication with the use of OPCs for sizing of atmospheric particles is that the instrument response varies with particle shape and composition. In particular, OPC response functions are nonmonotonic over some ranges of particle size, and particle light absorption (e.g., as occurs for soot-containing particles) leads to a flattened response curve that results in loss of size resolution. Nevertheless, OPCs are useful for obtaining number and size information for particles larger than 300 nm, where DMA techniques begin to have greater uncertainties. In particular, OPCs are the only currently commercially available instruments for sizing of supermicron particles, including cloud particles. The development of aerodynamic particle sizers (APS; e.g., Griffiths et al. 1984), which use particle inertia to size particles larger than about 500 nm, provided an additional measurement to constrain the aerosol size distribution up to about 20 μm. However, the APS response depends on particle shape and density. Hence inversion techniques are needed to harmonize the various methods for measuring particle size (e.g., Twomey 1963; Hand and Kreidenweis 2002; Kandlikar and Ramachandran 1999; Hansen 2000).

c. Particle mass concentrations and chemical composition

At the time of Aitken (late 1800s) and into the early 1900s, chemical measurements of precipitation and collected particles indicated that sulfate salts, which we now understand arise largely from oxidation of SO₂ and other sulfur gases from anthropogenic and natural sources, were a major component of the aerosol. Particles were conceptualized as internal mixtures composed of hygroscopic and insoluble components. This picture emerged from analyzing the change in size upon humidification of the aerosol, finding that the growth factor is less than what is expected for pure salts (Junge 1953). Junge’s work led to the description of the aerosol as a model salt (e.g., ammonium sulfate) internally mixed with an insoluble core. The soluble fraction ε is defined as the value that gives agreement between measured and modeled hygroscopic growth of the aerosol. The soluble fraction model was widely used for the next 50 years to represent the water uptake properties of the ambient aerosol. The soluble fraction as described by Junge represented the first single parameter parameterization of hygroscopic growth and CCN activity. Petters and Kreidenweis (2007, 2008, 2013) introduced the now widely used hygroscopicity parameter κ as an alternative single-parameter approach, showing that (approximately) 0 < κ < 1 for the range of constituents expected in the atmospheric aerosol, from nonhygroscopic (κ = 0) to the most hygroscopic salts. Thus, the hygroscopicity parameter serves as a useful scale for ranking relative affinity for water, and the volume of water in an aerosol at a particular relative humidity is proportional to the magnitude of the κ that describes it.

Remarkable advances in understanding of aerosol chemical composition have been made over the past century, primarily due to advances in analytical instrumentation that have enabled identification of molecular composition and that have also pushed the
boundaries of the smallest sizes of particles that can be analyzed. The establishment of monitoring networks in the United States (Solomon et al. 2014) and Europe (Tørseth et al. 2012), primarily associated with regulatory efforts in response to acid deposition, visibility, and human health impacts (see Wallington et al. 2019; Haupt et al. 2019b; Benjamin et al. 2019), have further provided long-term measurements of the seasonal and inter-annual variations in aerosol composition, and their relationships to meteorology and the local, regional, and global sources of atmospheric particulate matter.

Chow (1995) provides a review of methods that have been applied to the measurement of particulate mass concentrations, which often also afforded an opportunity to measure chemical speciation. Measurements in the late nineteenth century relied on gravitational settling of particles into “dustfall collectors” that were used to quantify airborne soluble species, but also collected confounding material (e.g., insects). Active sampling soon displaced these methods. Collection of particles by drawing known quantities of air through a filter and weighing the filter before and after exposure has long been a staple of atmospheric particulate matter (PM) mass measurements (Russell 1885). Advancements on this basic technique include the development of inlets with well-defined particle size cuts, so that the poorly defined early methodologies that reported “total suspended particulate matter” (TSP) could be refined into fractions such as PM$_{2.5}$ and PM$_{10}$—particulate matter having aerodynamic diameters less than 2.5 and 10 $\mu$m, respectively. These fractions were defined based on emerging information about the health impacts of particles (Heal et al. 2012), particularly the penetration efficiencies of various particle sizes into the human respiratory system (Cao et al. 2013). Filter-based collections of particles with subsequent chemical analyses that quantify many of the airborne components of the atmospheric aerosol continue to be mainstays of national and international aerosol monitoring networks [e.g., the U.S. Chemical Speciation Monitoring Network (CSN) and Interagency Monitoring of Protected Visual Environments (IMPROVE) programs; Solomon et al. 2014].

Impaction was an early method for fractioning the aerosol into size classes before sampling that exploited knowledge of the aerodynamics of particles to collect particles onto substrates. For example, wire impactors, flown on high-altitude aircraft, obtained the physical samples from which much of the early information on the physical and chemical characteristics of the stratospheric aerosol was derived (e.g., Pueschel 1996). By accelerating the sample flow, particles could be captured due to their inertia; careful design of the flow field, including the inlet characteristics and the collection surface, allows particles with diameters smaller than a designed size threshold to decelerate in the flow due to aerodynamic drag, and thereby follow fluid pathlines to bypass the collection surface. Marple (2004) provides a history of impactor design and use for aerosol sampling from 1860 to 1970, after which time advances in computational fluid dynamics permitted robust design of such devices for nearly any imagined application.

More recently, methods for aerosol speciation have undergone rapid development and broad application (Kolb and Worsnop 2012). Single-particle chemical analysis techniques (e.g., Bzdek et al. 2012; Prather et al. 1994) have enabled unprecedented insights into the composition, mixing states, and structure of ambient particulate matter, for both sub and supermicron fractions. These include laser ablation single-particle mass spectrometry (Prather et al. 1994; McKeown et al. 1991), transmission electron microscopy (TEM; Pöschl et al. 1994; Adachi and Buseck 2008), and scanning transmission X-ray microscopy (STXM; Russell et al. 2002; Takahama et al. 2007, 2010; Moffet et al. 2010). The development of the aerosol mass spectrometer (AMS) and its application to ambient aerosol monitoring (Jimenez et al. 2003) constituted a major breakthrough in the ability to observe, at high temporal resolution, changes in the nonrefractory composition of submicron particulate matter. Analysis of AMS data via techniques such as positive matrix factorization also revealed the possibility of identifying distinct aerosol types and sources from these data (Zhang et al. 2011). Figure 11-10 presents an overview of AMS measurements at Northern Hemisphere sites, demonstrating both the variability in mean composition as well as the dominance of organic carbon compounds in the submicron aerosol mass concentration at many locations. This picture contrasts with the very early conceptualizations of aerosol composition put forth by Junge and others, that pointed to the dominance of ionic species and in particular, of sulfate compounds. Several reasons for this discrepancy can be pointed to, including the difficulty in quantitatively measuring organic compounds as compared to ionic species. It is also worth noting the difference in particle sizes that these measurements capture, with sulfate mass preferentially found in the “droplet mode” at $\sim$700 nm (Meng and Seinfeld 1994), the size range where the measurement efficiency of the AMS begins to decrease. This droplet mode is formed by cycling of particles through cloud droplets, and uptake and conversion of gases into particulate matter in the aqueous phase (Wallington et al. 2019).

Perhaps one of the more surprising evolutions in understanding of atmospheric aerosol composition has been around the nature of sea spray aerosol (O’Dowd...
and de Leeuw 2007). Although early researchers recognized that salts other than sodium chloride were contained in sea spray particles (e.g., Simpson 1941), recognition of the role of the organic fraction, especially for ultrafine particles, has only become evident in recent years (e.g., Facchini et al. 2008; O'Dowd et al. 2004; Quinn et al. 2014). The organic fraction appears to be tied to biological activity (O'Dowd et al. 2004), and the atmospheric organic compounds are linked to seawater organic composition (Russell et al. 2010; Frossard et al. 2014).

Indeed, proper quantification of the mass concentrations in the ambient aerosol of organic compounds and elemental carbon is a long-standing problem in atmospheric aerosol research, and itself merits a separate review (e.g., Jacobson et al. 2000; McMurry 2000a). Comprehensive reviews by Russell (2014) and Wallington et al. (2019) discuss evolving insights into the formation routes and properties of secondary organic aerosol species, formed in the atmosphere via poorly understood precursors and pathways that include photochemistry, heterogeneous chemistry, and aqueous-phase chemistry. Most recently, attention has been focused on “brown carbon” (Laskin et al. 2015), organic carbon compounds that absorb preferentially at short wavelengths and thus may have underappreciated impacts on atmospheric photochemistry (Jacobson 1999) and climate (Feng et al. 2013). Brown carbon and soot are the primary shortwave-absorbing aerosol components in the atmosphere, and hence play key roles in Earth’s radiation budget (Lu et al. 2015; Bond et al. 2013; Boucher et al. 2013).

d. Aerosol hygroscopicity and water content

A dominant component of the atmospheric aerosol at moderate to high relative humidities is liquid water. An understanding that airborne particles comprise species that interact with atmospheric water vapor at subsaturated conditions is documented at least as far back as the work of Aitken, who showed that atmospheric visibility was decreased as relative humidity increased, even if the numbers of particles remained unchanged. Clues regarding the nature of salts in the aerosol were obtained by examining particles under a microscope under changing relative humidity, observing deliquescent and efflorescence phenomena, and relating these to known compounds. Some early investigations used spider webs to trap individual particles while minimizing contact with a substrate, and then exposed this system to varying relative humidities to measure changes in...
particle volume from which water contents could be deduced (e.g., Dessens 1949; Junge 1952; Twomey 1953, 1954; Fig. 11-11). Gravimetric methods, in which particles were deposited onto substrates and then weighed under variable humidity, were also employed (e.g., Speer et al. 1997; Winkler 1973) and continue to be applied in present-day studies. Winkler’s work explicitly noted the mixed-composition nature of the atmospheric aerosol, and the variability of these mixture compositions and their water contents depending on location (e.g., urban, rural, marine).

Single-particle analysis via electron microscopy would seem to be a suitable method to probe compositions of submicron particles. Indeed, we note that methods involving deposition of particles onto substrates for subsequent processing via microscopy were first employed decades ago and continue to be important in elucidating the nature of the ambient aerosol and its interactions with atmospheric water vapor under subsaturated conditions. In particular, environmental scanning electron microscopy (ESEM) has been used to probe reactions of deposited particles and subsequent changes in water uptake (Laskin et al. 2005; Moffet et al. 2010; Piens et al. 2016). Classic adsorption experiments applied to atmospherically relevant aerosol samples have shown that some atmospheric particles can take up multiple monolayers of water via this mechanism (e.g., Tang et al. 2016), leading to small but important aerosol water contents below 100% relative humidity. However, Junge (1953) already pointed out the inadequacy of microscopy methods; drawbacks include low throughput, loss of semivolatile material under vacuum, chemical and physical changes due to interaction with the electron beam, and lack of quantitative analysis. Further, all existing online and offline single-particle methods have limited ability to quantify the mass present in each particle (e.g., Bahadur et al. 2010). Nonetheless, single-particle analysis with transmission electron microscopy and STXM have provided some constraints on particle morphology, mixing state, and elemental composition (Adachi and Buseck 2008; Bzdek et al. 2012; Moffet et al. 2010). Chemical analyses from particles sampled with cascade impactors can extend into the sub-100-nm size range but suffer from potential artifacts associated with low sample volume, particle bounce, and substrate interference (Chow and Watson 2007). Nonisokinetic sampling, especially on moving platforms, may also bias results (Wendisch et al. 2004).

By the late 1980s, new methods had been developed to allow for substrate-free experimental determination of aerosol water content as a function of relative humidity for single particles and particle populations, and various treatments of the thermodynamics of the corresponding equilibria for use in modeling of visibility and climate effects were also being developed. A recent review of these developments, including summaries of...
observational data for various aerosol types, is provided by Kreidenweis and Asa-Awuku (2014) and, therefore, only some are briefly summarized here. The electrodynamic balance traps a supermicron-sized particle in an electric field and by measuring the voltage required to balance gravity, changes in mass can be inferred as the relative humidity within the trap is modified (Cohen et al. 1987). In a series of influential papers, Tang and coworkers (e.g., Tang 1997) applied this technique to a large range of single- and multi-component particles composed of species that may be present in the ambient aerosol. More recently, Reid and coworkers showed how optical tweezers can be used to trap single particles (Hopkins et al. 2004) and measure not only hygroscopic growth, but also other properties of interest to the atmospheric aerosol (Cai et al. 2015; Power and Reid 2014).

A seminal development for probing water content in single, submicron particles was the humidified tandem differential mobility analyzer (H-TDMA; Rader and McMurry 1986), an extremely sensitive technique that is capable of distinguishing the addition of a few monolayers of water. The H-TDMA concept has subsequently been reproduced by researchers around the world (Duplissy et al. 2009; Massling et al. 2011), extended to work at relative humidity approaching water saturation (Hennig et al. 2005; Suda and Petters 2013), and used extensively in both laboratory and field settings to probe hygroscopicity as a function of particle size (e.g., Hämeri et al. 2001), including for nanoparticles < 50 nm (Russell and Ming 2002; H. Chen et al. 2018; Biskos et al. 2006; Kim et al. 2016).

More recently, aerosol mass spectrometry has advanced to provide rich information about composition extending to the sub-100-nm regime (Jimenez et al. 2009; Lawler et al. 2014). Using mass spectrometric methods combined with measurement of water contents, the influence of organic compounds toward ambient CCN concentrations has become much better understood (Chang et al. 2010; Gunthe et al. 2009; Levin et al. 2014; Mei et al. 2013). Combining atmospheric studies with results from fundamental laboratory studies (Cappa et al. 2011; King et al. 2007; Lambe et al. 2011; Massoli et al. 2010; Petters et al. 2017; Prenni et al. 2007; Rickards et al. 2013; Ruehl et al. 2016; Suda et al. 2014; VanReken et al. 2005; Wex et al. 2007) and theoretical considerations (Nakao 2017; Petters et al. 2016; Petters et al. 2009a) firmly constrains the contribution of organic compounds to CCN activity to 0 < \( \kappa_{\text{org}} \) < 0.3, where \( \kappa_{\text{org}} \) is the effective hygroscopicity parameter (Petters and Kreidenweis 2007) of the organic fraction. Furthermore, \( \kappa_{\text{org}} \) tends to increase with oxidation and aerosol age (e.g., Jimenez et al. 2009; Lambe et al. 2011; Massoli et al. 2010), although this is not always the case (Cerully et al. 2015). For example, modeling simultaneous dilution and chemistry during evolution of smoke plumes demonstrates that “aging” includes microphysical changes as well as changes in composition (Bian et al. 2017), both affecting CCN activities.

Figure 11-12 demonstrates the range of subsaturated hygroscopic growth that has been measured for particles composed of proxy pure compounds. We note that these equilibrium water contents, due to the curvature effect, vary with dry particle size for small (submicron) particles, with small particles being more concentrated than large particles in equilibrium with the same ambient relative humidity. In general, for a given particle size, salts (and especially sodium salts) exhibit the largest water contents across the range of ambient relative humidities. As noted above, much attention has been given

![Figure 11-12. Aerosol water contents, expressed as the ratio of the volume of water to the volume of dry particle, for pure compounds found in the atmospheric aerosol. Solid lines are for stable solution states, whereas dotted lines are for metastable (supersaturated) solutions. Calculations were performed at 298 K using the E-AIM model (Wexler and Clegg 2002). [Figure from Kreidenweis et al. (2008). © IOP Publishing.](image-url)
in recent years to measuring and modeling the hygroscopicities of various organic carbon components of the atmospheric aerosol (e.g., Petters et al. 2016; Petters et al. 2009a; Rickards et al. 2013 for data compilations). While in general the water contents associated with organic species are modest in comparison with ionic compounds, solution nonidealities may change this picture. Figure 11-12 also shows deliquescence and efflorescence behaviors, which have been documented to occur in the ambient aerosol. Kreidenweis and Asa-Awuku (2014) have provided recent compilations of observational data on atmospheric aerosol hygroscopicity for various source types, locations, and seasons.

e. NPF

Some of the most striking advances in understanding of atmospheric aerosols have occurred in the detection of NPF events, which are now recognized to play a major role in controlling atmospheric particle number concentrations, and in the elucidation of the chemical species and mechanisms involved. Interestingly, intimations that photochemical processes can affect particle number concentrations can be found in the writings of Aitken prior to 1900. However, consistent detection of the presence of nucleation mode particles (∼3–20 nm in diameter) and observations of their evolution in size space were not possible until development of appropriate, calibrated counting and sizing systems (e.g., Kulmala et al. 2004; McMurry 2000a). Hence, observations of bursts of nucleation-mode number concentrations were not widely reported until the 1990s, and the first observations that were proposed to be due to NPF were made in cloud outflows (e.g., Perry and Hobbs 1994) or in the free troposphere, especially in the outflow of the intertropical convergence zone (e.g., Clarke et al. 1999). Those observations, at cold temperatures and high relative humidities, were consistent with expectations for the conditions conducive to the nucleation of sulfuric acid–water particles. To our knowledge, Doyle (1961) was the first to propose that new particles could be formed in the atmosphere via binary homogeneous nucleation from these vapors, and nucleation in the sulfuric acid–water system was subsequently studied extensively in the laboratory and theoretically (Wyslouzil and Wölk 2016). Interest in nucleation as a fundamental atmospheric process shaping the atmospheric aerosol intensified as many more instances of NPF in the ambient atmosphere were documented, including in urban boundary layers (e.g., Woo et al. 2001), where it had previously been assumed that scavenging of precursor vapors would be too rapid to allow the buildup of the required saturation ratios to support NPF. It is now understood that NPF is ubiquitous in the atmosphere and can occur by a number of mechanisms that also involve organic and nitrogen species (Kulmala et al. 2013; Lee et al. 2019).

f. Theory and modeling

Understanding of the factors shaping the atmospheric aerosol evolved rapidly. According to Junge’s analysis, the most prominent mode of the number size distribution is generally between 10 and 100 nm. For sizes exceeding 100 nm, he found that the number decreased as a power law with particle size, corresponding to a linear distribution when graphed in log–log space (Fig. 11-13). The idea that the atmospheric aerosol reached an equilibrium with respect to the processes of production, coagulation, and sedimentation led Friedlander (1960a,b) to develop governing equations...
for the kinetics of these processes, as well as to propose a similarity transform and the theory of self-preserving size distributions, arriving at a dimensionally estimated slope for the large particle size range close to that proposed by Junge. Clark and Whitby (1967) tested the proposed relationships against experimental data for ambient aerosols collected using a continuous-flow system, comprising an electrical particle counter, a condensation nuclei counter, and an optical counter to cover the complete size range from \( \sim 2 \text{ nm} \) to \( 6 \mu \text{m} \) in diameter. They confirmed the Junge power-law distribution for particles with diameters larger than \( \sim 100 \text{ nm} \), but noted that “below this size, the magnitude and shape of the distribution function was highly dependent on local source and weather conditions.” Subsequent work led Whitby and coworkers (Whitby et al. 1972) to introduce the concept of the multimodal atmospheric aerosol, as well as to suggest that a lognormal function could be used to fit the data (Fig. 11-14), as also described by Jaenicke and Davies (1976). The magnitude and composition of the “accumulation range” in this figure is key to many of the aerosol–climate effects mentioned herein, as these particles have relatively long lifetimes in the atmosphere due to limited removal mechanisms, and thus can exert influence on regional to global scales; further, this size range interacts most efficiently, per unit mass of particulate matter, with solar radiation, and also generally constitutes the predominant number fraction of CCN. Gelbard and Seinfeld (1979) developed the discrete-continuous general dynamic equation for aerosol size distribution dynamics. The ideas put forth in these and subsequent publications continue to form the foundation of how the atmospheric aerosol number, surface area, and mass distributions and their microphysical transformations are conceptualized today.

![Diagram of atmospheric aerosol surface area distribution](image)
Due to space limitations, the discussion here cannot adequately cover the myriad advances that have occurred in the understanding of atmospheric particulate matter that have been driven by research in other areas; many of these are covered in companion articles, which we mention briefly here. Wallington et al. (2019) discuss the history of tropospheric and stratospheric chemistry research. Concerns regarding the health effects of degraded air quality motivated early studies of the sources and fates of pollutants, with atmospheric particulate matter, particularly soot and sulfuric acid, as obvious components of sulfurous, “London-type” smog and of acid rain. The discovery in the 1950s of the chemistry leading to photochemical smog showed how particulate matter could be created via gas-phase chemistry involving anthropogenic and biogenic precursors, and spurred intensive studies of the formation pathways and chemical composition of secondary organic aerosol (SOA); characterization of SOA via new analytical techniques and furthering understanding of its role in climate and health continue to be active areas of research. Fundamental understanding gained from laboratory and field studies is tested via models of chemical reaction mechanisms, and the interactions between chemistry and the atmosphere are tested via transport, transformation, and aerosol–chemistry–climate models. Indeed, Wallington et al. (2019) note that coupling of chemistry and climate change, and human health impacts, are current strong drivers of atmospheric chemistry research directions. Haupt et al. (2019b) and Benjamin et al. (2019) discuss the development of air pollution meteorology and air quality forecasting models, which range in scale from local (e.g., plume dispersion models) to global. Randall et al. (2019) describe the historical development of Earth system models, for which treatments of atmospheric chemistry sources, transformations and sinks have also been developed. As discussed briefly below, regional and global models have been used to predict past, present-day, and future global atmospheric aerosol fields for a range of applications that include air quality/health and weather/climate interactions. Further, information on atmospheric aerosol fields is needed to drive models that seek to simulate aerosol–cloud–precipitation interactions via sophisticated multimoment cloud microphysical schemes (see Randall et al. 2019, for descriptions of the evolution of cloud microphysics models). This requirement has led to new modeling approaches that seek to simulate not only aerosol mass concentrations, but also number size distributions, requiring representation of aerosol microphysical processes (e.g., Adams and Seinfeld 2002; Stier et al. 2005).

### 3. Cloud condensation nuclei

The specific subset of the atmospheric aerosol that serves as sites for the condensation of water under conditions found in the atmosphere, leading to formation of water droplets, is known as CCN. The scientific exploration of condensation nuclei began with John Aitken (1839–1919). His expansion chamber experiments showed that water vapor always condenses onto preexisting nuclei and that some particles are more efficient than others at nucleating water droplets (Podzimek 1989). In the 1920–30s, the microphysics of the formation of cloud droplets and precipitation drops became better understood. Hilding Köhler (1888–1982) developed the theory for equilibrium hygroscopic growth and droplet activation (Köhler 1936). Köhler’s calculations first constrained the characteristic dimension of the nuclei to \( \sim 100 \text{ nm} \). Köhler (1936) juxtaposed observations of chlorine contents in hoar frost, fog and clouds, and rainwater, and found concentrations to be similar in all. This finding suggested to him that raindrops must form from combinations of millions of cloud droplets, and not by growth of cloud drops into raindrops via condensation of pure water from the vapor phase, which would result in strong dilution as drop sizes increased.

Figure 11-15 illustrates classical Köhler theory. Liquid equilibrium states of an illustrative single-component, soluble, spherical dry particle are shown as a function of the saturation ratio \( S \). For subsaturated states \( (S < 1) \), the aerosol takes up water leading to an increase in the particle diameter. The water uptake is characterized by the ratio of the wet to dry particle diameters.
commonly referred to as the hygroscopic growth factor (gf). In the case of absorption, subsaturated hygroscopic growth is driven by the change in free energy of mixing of water and solutes, which is approximated to first order by Raoult’s law. As illustrated in Fig. 11-15, actual growth factors are less than the Raoult’s law prediction due to the energy cost of forming a surface (the Kelvin effect: Thomson 1871; Sir William Thomson, Lord Kelvin, 1824–1907). As a consequence, equilibrium growth factors exist at supersaturated states, where supersaturation $s$ is defined as $s = (S - 1) \times 100\%$. If the ambient saturation ratio $S$ exceeds a critical value $s_c$, no equilibria exist. A particle embedded in an environment where the environmental $s$ exceeds $s_c$ serves as a CCN and grows by diffusion into a cloud droplet. The $s_c$ of a particle is controlled by the amount and type of absorbed or absorbed water by the solutes, the amount of trace gases (e.g., HNO$_3$ and organic gases) present in the environment, and the tension between the air/liquid interface. In principle, each particle has a unique equilibrium curve with water vapor that depends on its size, composition, and morphology. Furthermore, each particle’s $s_c$ may vary with the concentration of atmospheric trace gases. As a practical simplification, particles can be grouped into the cumulative number of particles that have $s_c \leq s$.

Starting with Köhler theory, numerical cloud models were developed by Howell (1949) and Squires (1952). Their work established the well-known supersaturation balance equation that predicts a maximum value a few tens of meters above thermodynamic cloud base. The first numerical cloud models were motivated by the desire to understand the formation of precipitation. Using model simulations, Howell inferred that atmospheric supersaturations will range between 0.1% and 1% supersaturation, that droplet number is controlled by the maximum of the vertical supersaturation profile, that drop number is only weakly sensitive to nuclei concentration, and that additional processes are needed to explain the appearance of larger droplets that initiate rain. One limitation in validating the numerical models was an absence of direct measurements of particles serving as cloud droplet nuclei.

The first successful direct measurements of cumulative CCN activation spectra were made by Wieland (1956) using a mixing thermal gradient diffusion chamber. Prior attempts at direct measurements had failed due to the inability to precisely control the supersaturation inside the measurement apparatus. Wieland’s data showed a strong increase of the available CCN with supersaturation. Concentrations measured in a wide variety of airmasses exposed to supersaturations between 0.1% and 1% ranged from 50 to 5000 cm$^{-3}$. Prior measurements by Diem (1948) had also constrained cloud droplet number concentrations, which ranged between 260 and 1000 cm$^{-3}$ for different of cloud types, including stratus and cumulus. Wieland’s measurements confirmed typical maximum atmospheric water supersaturations between 0.2% and 1%, as suggested by the numerical cloud models.

Twomey (1959a,b) first sought formal closure between measured CCN supersaturation spectra and observed cloud droplet concentrations. Twomey represented measurements using a power-law relationship of the form

$$N_{CCN} = C s^k,$$

where $N_{CCN}$ denotes the number of CCN measured using the CCN instrument, $s$ denotes the supersaturation, and $C$ and $k$ are fitted coefficients. Twomey then derived approximate upper and lower bound solutions of the supersaturation balance equation that readily provide cloud droplet number concentrations $N_d$ as functions of $C$, $k$, and updraft velocity $w$:

$$N_d \approx C^{2(k+2)} \left[ \frac{\psi(T, p)\beta^{3/2}}{kB \left(\frac{T}{T_0}\right)^{3/2}} \right]^{k(k+2)},$$

where $\psi(T, p)$ is a coefficient that subsumes the supersaturation generation rate and droplet growth rate, and $B(a, b)$ is the beta function. Studies of CCN to cloud droplet number concentration closure suggest that Twomey’s parameterization is suitable for predicting CCN–cloud drop relationships under conditions where $C$, $k$, and $w$ are directly measured (Fig. 11-16; Snider and Brenguier 2000; Twomey and Warner 1967; Yum et al. 1998).

As will be discussed further below, Twomey’s parameterization introduced a powerful means to estimate how changes in CCN due to anthropogenic activity lead to changes in cloud droplet number. Furthermore, Twomey’s analysis again suggested that in-cloud supersaturation will rarely exceed 1%. A significant fraction of the total condensation nuclei (CN) pool can only be activated at $s > 1\%$. Because atmospheric supersaturations were generally considered to be $<1\%$, CCN number concentrations for $s \leq 1\%$ represent only a subset of the CN, and the two are decorrelated in some environments. Thus, CCN emerged as a conceptually distinct subset of the CN. How this subset was related to the total condensation nuclei pool remained unclear.

Köhler’s original theory predicted that the activation supersaturation depends on particle size and chemical composition. However, quantitatively linking aerosol size and chemical composition to predict CCN spectra proved to be a formidable challenge that remains a topic of active research. Successful closure depends on precise representation of the size-resolved aerosol chemical
composition, precise size distribution data, precise CCN concentration and supersaturation characterization inside CCN instruments, and a correct theory linking size, composition, and activation supersaturation, the latter depending also on the updraft velocity (i.e., competing kinetics of relevant processes). Some of the main developments in each of these areas are now summarized.

a. Theory

Köhler theory was refined and extended to account for nonideal solutions (McDonald 1953), to treat internal mixtures (Fitzgerald 1973), to treat activation of insoluble (Fletcher 1958; Henson 2007; Laaksonen and Malila 2016; Mahata and Alofs 1975; Sorjamaa and Laaksonen 2007) and sparingly soluble (Shulman et al. 1996) nuclei, to account for the presence of water soluble gases (Kulmala et al. 1997), to account for surface tension effects (Li et al. 1998; Nozière et al. 2014; Petters and Petters 2016; Ruehl et al. 2016; Sorjamaa et al. 2004), and to account for cocondensation effects during adiabatic ascent (Topping et al. 2013; Topping and McFiggans 2012). Several studies in the 1990s and early 2000s demonstrated the important contribution of organic compounds in modulating CCN activity (Cruz and Pandis 1998; Novakov and Penner 1993; Prenni et al. 2007; Raymond and Pandis 2002; VanReken et al. 2005). These studies sparked interest in how to parameterize and predict the contribution of organic compounds to CCN activity. Due to the large number of compounds and lack of complete speciation, single parameter descriptions of CCN activity within the framework of Köhler theory, which originated between the 1950 and 1970s (Fitzgerald 1973; Junge 1953) were reworked to describe complex organic mixtures (Petters and Kreidenweis 2007, 2008, 2013; Rissler et al. 2006; Wex et al. 2007). The single parameter models were first used to characterize the contribution of the organic fraction to CCN activation, without the need to attribute the activity to specific physicochemical properties of individual components. To predict the CCN activity of organic compounds with known molecular structure, thermodynamic models based on functional group contribution methods have been further developed to integrate with Köhler theory (Ming and Russell 2004; Petters et al. 2016; Topping et al. 2016; Topping et al. 2005). The foregoing body of work applies to describing the equilibrium Köhler curve. In addition, potential kinetic effects hindering the particle from being in equilibrium have been explored. These include retarded condensation rates due to organic films (Chuang et al. 1997; Nenes et al. 2001b), retarded water uptake due to slow dissolution kinetics (Asa-Awuku and Nenes 2007), and retarded rates to establish equilibrium surface tension on the droplet (Nozière et al. 2014).

b. Instrumentation

After Wieland’s successful series of experiments established the feasibility of controlling supersaturation in
order to achieve direct measurement of cumulative CCN spectra, CCN measurement techniques were refined and improved (e.g., Hallett 1983). Wieland's chamber was a turbulent static diffusion chamber. Twomey's instrument was similar in design but used a chemical diffusion method. The community then settled on the static diffusion cloud chamber, which uses a stably stratified chamber with a thermal gradient between two water saturated plates to produce supersaturated conditions. In all of these methods, the maximum supersaturation forms near the midpoint of the two plates. The initial method for quantifying CCN concentrations involved activating the droplets and letting them settle on paper, followed by counting of the droplets. Alternatively, the droplets could be illuminated, photographed, and bright spots on the image counted manually. Radke and Hobbs (1969) used bulk light scattering to determine CCN concentrations via a calibration between peak light scattering and directly imaged droplets. This technique was widely adopted in subsequent similar static diffusion designs (Delene and Deshler 2000; Lala and Jiusto 1977).

In parallel, continuous-flow instruments were developed (Chuang et al. 2000; Fukuta and Saxena 1979; Hudson and Squires 1976; Ji et al. 1998), each with a unique principle of operation. The Hudson and Squires CCN spectrometer was used extensively for airborne observations for several decades. Chuang et al. (2000) introduced a continuous-flow CCN instrument based on a new design principle. The design was further refined by Roberts and Nenes (2005) and a commercial spinoff of that instrument design led by Droplet Measurement Technologies (DMT) has resulted in an explosive growth of CCN measurements over the past decade. The DMT instrument is a continuous-flow streamwise-gradient instrument. The aerosol is focused in the centerline of a tube that is wetted with water. A stable centerline supersaturation forms in the upper third of the instrument. Droplets that activate grow to supermicron size and are detected using an optical particle counter. A key advantage of this particular instrument is the ability to acquire CCN concentrations at constant supersaturation and 1-Hz frequency. Precision and time resolution are vastly improved compared to the bulk light scattering technique with static diffusion chambers.

Diffusion theory was used to calculate the supersaturation in static thermal gradient CCN instruments as a function of temperature gradient. Initial estimates were biased and high due to neglecting the Soret and Dufour effects, as clarified by Katz and Mirabel (1975). However, nominal supersaturations determined from the activation of monodisperse ammonium sulfate and sodium chloride remain in disagreement (Bilde and Svenningsson 2004; Snider et al. 2006). Calculation of supersaturation in the streamwise-gradient design (Roberts and Nenes 2005) requires more complex models that include treatments of heat transfer, fluid flow, and diffusional droplet growth (Lance et al. 2006; Raatikainen et al. 2012; Roberts and Nenes 2005; Snider et al. 2010). As with the static chambers, measured and modeled supersaturations were not easily reconciled (Lance et al. 2006). Due to the discrepancy between modeled and observed supersaturations in instruments of both the static thermal gradient design and streamwise-gradient design (Rose et al. 2008; Snider et al. 2006), the accepted methodology is to calibrate the instrument supersaturation using monodisperse particles of known composition, usually ammonium sulfate. This is significant, as calibration requires calculation of the supersaturation through Köhler theory. The reported instrument supersaturations are therefore relative to an equivalent salt. Further, dried salt particles are not spherical, leading to additional uncertainties regarding the precise mass of salt that should be used in Köhler theory calculations. Until there is closure between supersaturation computed from theory and instrument characteristics alone, and supersaturation predicted by Köhler theory for the test particles, the absolute verification of Köhler theory through this CCN measurement remains unsolved. This discrepancy complicates attempts at closure between particle chemical composition, Köhler theory, and measured CCN concentrations in the atmosphere.

c. Aerosol–CCN–cloud droplet closure

Fitzgerald (1973) reported poor agreement between modeled and predicted CCN spectra, citing possible instrument errors. Fitzgerald also showed how Junge’s soluble fraction, which is inferred from growth factor data, is related to the relationship between critical supersaturation and dry diameter through Köhler theory. Bigg (1986) first explicitly suggested that observational data did not support aerosol-to-CCN closure. Bigg compared 4 years of CCN measurements made with a static gradient diffusion counter and size distributions obtained with an optical particle counter and a diffusion battery coupled with a Pollack counter (Gras and Ayers 1983). Chemical composition was constrained via electron microscopy. Predicted CCN from size and chemical composition exceeded observed CCN by a factor of 3–10. Bigg attributed the difference to organic compounds that delay drop formation. A number of aerosol-to-CCN closure studies in the early 2000s yielded significantly better closure, although predicted CCN concentrations...
from Köhler theory frequently exceeded measured values by up to 30% (Cantrell et al. 2001; Covert et al. 1998; Snider et al. 2003; VanReken et al. 2003; Zhou et al. 2001). Aerosol–cloud droplet closure studies also encountered difficulties, including necessitating the assumption of low mass accommodation coefficients (e.g., Meskhidze et al. 2005). Subsequently, research on the aerosol-to-CCN closure problem intensified. Reports of successful closure are not uncommon (e.g., Fig. 11-16; Fountoukis et al. 2007; Mochida et al. 2010; Sanchez et al. 2018; Vestin et al. 2007). However, two synthesis studies from this decade demonstrate that the closure problem has not been solved entirely (e.g., Kammermann et al. 2010; Whitehead et al. 2014). Successful aerosol-to-CCN closure necessitates accurate measurement of particle size/mass (e.g., Sanchez et al. 2016) as well as composition. Nonetheless, starting with Bigg (1986), most authors cite the presence of organic compounds to be associated with lack of aerosol-to-CCN closure. A recent example is Ovadnevaite et al. (2017), who report lack of closure only under conditions when the aerosol is composed of a mixture of organic and inorganic compounds. With respect to the ability to predict droplet number concentrations from information about the aerosol, the feedbacks with cloud dynamical processes complicate aerosol-to-cloud droplet closure. For example, entrainment negates the applicability of the assumption of adiabatic updrafts, requiring stratification of data and/or explicit representation of entrainment in the model to make appropriate comparisons (e.g., Hudson and Yum 2002; Conant et al. 2004; Sanchez et al. 2017). Further, microphysical processes, including the formation of drizzle, affect droplet number concentrations and limit the ability to use observations to directly test for aerosol–cloud droplet closure (e.g., Wood et al. 2012).

d. Composition

As shown by the aerosol-to-CCN closure studies, the critical size range at which atmospheric particles become CCN active under atmospheric supersaturations between 0.1% and 1% is between 150 and 30 nm. The exact cutoff diameter depends on the particle chemical composition and imposed supersaturation. Composition constraints on aerosol-to-CCN closure therefore need to target composition in the appropriate size range, including sub-100-nm particles. However, as discussed above in the sections addressing aerosol composition and hygroscopicity, obtaining accurate composition measurements, particularly in the sub-100-nm size range, has been challenging. As a consequence, composition constraints on early aerosol-to-CCN closure studies were weak. For this reason, Junge’s idea to use aerosol water uptake measurements as a proxy measure for composition has been widely embraced. However, the constraint from growth factor measurements remains imperfect. Further, aerosol hygroscopicity represents an aggregate property, and unique mapping from hygroscopicity to composition is not possible.

4. Ice nucleating particles

a. Historical development

Bergeron (1935) and Findeisen (1938) determined that the appearance of a small number of ice crystals among a cloud of liquid droplets leads to rapid growth of the ice particle by vapor transfer from the supercooled liquid to ice surface; the grown crystal then falls and grows further by riming and aggregation. At the time it was not understood whether ice crystals originated from spontaneous freezing of drops or whether some nucleus was required to initiate the first ice crystals (Simpson 1941). Initially these hypothetical nuclei were referred to as sublimation nuclei, particles that initiate ice directly from the vapor phase, just above ice saturation. Experiments in the 1940s with expansion chambers showed that the nuclei active in the mixed-phase cloud regime are insoluble, wetted with water, and susceptible to deactivation by heat and/or freeze–thaw cycling (Mason 1950). Experiments also showed that significant ice supersaturation or water saturation was required to initiate freezing. Therefore, the terms freezing nuclei (Mason 1950) or ice nuclei (Schaefer 1946) were used to describe these particles. Langham and Mason (1958) convincingly summarized various measurements to show the temperature at which pure water drops freeze homogeneously is ∼−38°C. This temperature delineates the warmer mixed-phase cloud from colder cirrus cloud regimes. In parallel, homogeneous and heterogeneous nucleation theories were developed by Vollmer, Turnbull, Fletcher, and coworkers (Fletcher 1958, 1959; Turnbull 1950; Turnbull and Fisher 1949; Vollmer and Weber 1926). Nucleation theory provided a semi-quantitative conceptual framework that identified the main factors influencing ice nucleation: ice supersaturation, degree of supercooling, the need to form a critical germ of a certain size to favor ice formation, and the contribution of random fluctuations and surface catalytic effects that control the germ size distribution. Experiments by Vali and Stansbury (1966) provided a means to disentangle the interplay between stochastic and deterministic effects in heterogeneous freezing. These experiments demonstrated that specific active sites that catalyze freezing are present on particle surfaces. The temperature at which a site initiates freezing...
fluctuates randomly by 1°–2°C when the same nucleus is undergoing freeze–thaw cycles (e.g., Shaw et al. 2005). This random variability (or kinetic rate effect) on freezing temperature is much smaller than the increase in nucleation rate with temperature. Therefore, it is possible to conceptualize a quasi-deterministic INP that initiates freezing at some specified supercooling temperature, as shown in Fig. 11-6.

A number of apparatuses provided the first measurements of the number of INP present in the atmosphere (Bigg 1957; Georgii 1959; Langer 1973). Georgii (1959) showed a close correlation between freezing nuclei concentration and number concentration of particles > 500 nm. Georgii (1959) and others (Isono and Tanaka 1966; Langer et al. 1979) also showed an influence of meteorology, specifically noting a sharp increase in airborne ice nuclei during and after storms. Rau (1954) showed a dependence on airmass. Fletcher (1962) and Bigg and Stevenson (1970) provided the first parameterization of atmospheric INP concentrations, based on correlation of INP concentrations from different sites with temperature.

Laboratory studies, atmospheric INP measurements, and electron microscopy on snow crystals identified mineral dust as an important source of atmospheric INPs (e.g., Kumai 1951; Mossop 1963; Roberts and Hallett 1968; Schaefer 1950). The experiments by Schaefer (1946) had shown that artificial clouds formed on laboratory aerosol did not glaciate even at supercooled temperatures. However, rapid ice formation occurred when injecting the cloud with dry ice. In response, Vonnegut (1947) searched through X-ray crystallographic data, hypothesizing that effective ice nucleating substances would have crystalline structures similar to ice, and identified silver iodide particles as INP that are active at temperatures warmer than −10°C, a hypothesis confirmed by laboratory and field experiments. Since then, a number of substances that are more or less efficient in nucleating ice have been described. Schnell and Vali (1972, 1975) identified ice nucleators in decomposing leaf litter and marine plankton. This work led to the discovery of ice nucleating bacteria (Pseudomonas syringae and Erwinia herbicola; Vali et al. 1976; Lindow et al. 1978; Levin et al. 1980; Yankofsky et al. 1981), which remain the most efficient ice nucleators known. The ice nucleation activity of bacteria is due to membrane bound proteins, which catalyze the critical cluster formation. The conformation of the ice nucleation (IN) active protein influences the IN temperature (Morris et al. 2004). Other less efficient biological ice nucleators, including algae, fungal spores, lichen, pollen, and water-soluble macromolecules have been identified (e.g., Fröhlich-Nowoisky et al. 2016; Pummer et al. 2015 and references therein). Several organic substances were also identified as more or less efficient INPs, including alcohol monolayers, amino acids, and black carbon (DeMott et al. 1999; Fukuta 1966; Gavish et al. 1990, 1992; Schill et al. 2016; Mahrt et al. 2018). The extent to which biological particles and/or carbonaceous substances contribute to the number concentration of atmospheric INP and compete with mineral dust sources remains an area of active research [e.g., as outlined in the reviews by Hoose and Möhler (2012) and Murray et al. (2012)].

During the last few decades, significant progress in laboratory and field measurements of INPs and cloud ice properties has been made. However, predicting whether a particle will nucleate ice from presently measured physical and chemical characteristics remains elusive. Quantitative theories that capture cloud ice evolution based on INP spectra are also lacking. The main developments in these areas are now summarized.

b. Theory

Over the past century, distinct mechanisms by which particles can initiate ice formation in the atmosphere have been discovered and described: homogeneous nucleation of pure water from dilute solutions (cloud and raindrops), homogeneous freezing of concentrated solutions (haze drops), and freezing nucleation (direct vapor-to-ice transitions). Appropriate nomenclature and process descriptions are given by Vali and coworkers (Vali 1985; Vali et al. 2015). Not all of these mechanisms are equally well understood. The current state of knowledge on these mechanisms is briefly described. Definitions of each process are taken verbatim from Vali et al. (2015).

1) Homogeneous Freezing

Homogeneous nucleation refers to “ice nucleation within a body of supercooled liquid without any foreign substance aiding the process” (Vali et al. 2015). Experimental data on the homogenous freezing limit are given by Langham and Mason (1958). Supercooled pure water drops nucleate homogeneously between −38° and −33°C for drop diameters ranging between 10 μm and 1 mm, respectively. A complete theoretical description of homogeneous nucleation of supercooled water has been given by Jeffery and Austin (1997). Solutes depress the melting point. Koop et al. (2000) showed that water activity, and by extension relative humidity with respect to ice, is the controlling variable for homogeneous nucleation of haze particles. The “Koop” freezing line is valid at temperatures less than the homogeneous limit for water and presents an
upper limit of ice saturation for which the upper free troposphere can maintain cirrus-free conditions.

2) Condensation/Immersion Freezing Nucleation

Freezing nucleation refers to "ice nucleation within a body of supercooled liquid ascribed to the presence of an INP, or equivalent. Immersion freezing refers to ice nucleation initiated by an INP, or equivalent, located within the body of liquid. Condensation freezing is defined as taking place when freezing is initiated concurrently with the initial formation of liquid on a CCN at temperatures below the melting point of ice" (Vali et al. 2015). These mechanisms are active in mixed-phase clouds and are the mechanisms probed by the measurement techniques used to obtain the data in Fig. 11-6. The data in Fig. 11-6 also summarize the range of observations of immersion mode INP concentrations identified in the atmosphere. Significant worldwide research focus on immersion freezing nuclei, especially in the last decade, has produced the most extensive data coverage for this mode.

3) Deposition Freezing Nucleation

Deposition freezing nucleation refers to "ice nucleation from supersaturated vapor on an INP or equivalent without prior formation of liquid" (Vali et al. 2015). Measurements of dust INP in the deposition mode suggest that for the same substance fewer particles nucleate in the deposition mode than the immersion mode (e.g., Sullivan et al. 2010a). The presence of deposition freezing nuclei has been shown to be most critical in the upper free troposphere at temperatures below the homogeneous freezing temperature of pure water. If such nuclei are absent, the upper free troposphere can remain supersaturated and cloud free for humidities below the Koop limit (Krämer et al. 2009). If present, their number (and freezing relative humidity) exerts significant control on cirrus cloud microphysics (Barahona and Nenes 2009; DeMott et al. 1994). Recently, Marcolli (2014) suggested that this mode is not present in the atmosphere, and rather freezing of liquid in pores is responsible for observations previously attributed to this mechanism.

4) Contact Freezing Nucleation

Contact freezing is "initiated by an INP, or equivalent, at the air–water interface as the INP comes into contact with the liquid, or forms at an air–liquid–particle triple interface" (Vali et al. 2015). Contact with the interface can be made from within the liquid or from collision of a drop with the air. Relatively few recent studies have focused on contact freezing (e.g., Durant and Shaw 2005; Fornea et al. 2009; Gurganus et al. 2014; Niehaus et al. 2014; Shaw et al. 2005). Niehaus et al. (2014) estimate that $10^3$–$10^5$ collisions are required to initiate a contact freezing nucleation event by mineral dust between $-15^\circ$ and $-20^\circ$C. Although it is difficult to estimate the importance of the pathway, one would expect that a significant fraction of large dust particles (diameters > 200 nm) serves as CCN and is immersed in cloud water and hence absent as interstitial particles. Collisions of interstitial particles with cloud droplets are thus likely too rare for this pathway to compete with condensation/immersion freezing. Definitive evidence to rule out this pathway, however, will require integration of the collection kernel with realistic dust particle number, cloud droplet concentrations, size, and contact freezing efficiencies (e.g., building on the work of Young 1974).

Heterogeneous freezing nucleation is thought to initiate on preferential sites, which have been proposed to be crystal defects, pores, cracks, and chemically modified surfaces (Fletcher 1969; Gorbunov et al. 2001; Knight 1979; Marcolli 2014; Sear 2011; Vonnegut 1947). However, the exact nature of these sites remains unclear due to the difficulty in visualizing the surface at the time and length scale of the nucleation event; simulated critical cluster diameters for homogenous ice nucleation range between 2.4 and 8 nm (Sanz et al. 2013). Recent advances in environmental scanning electron microscopy approach the required scale and confirm the concept of active sites (Kiselev et al. 2017; Wang et al. 2016). For example, Kiselev et al. (2017) obtained imagery that shows how nucleation proceeds on steps, cracks, and inside cavities on potassium feldspar (Fig. 11-17). Associated molecular dynamics simulations support the idea of chemical priming by the surface through the addition of hydroxyl groups. This lends support to descriptions of the ice nucleation activity of single substances using stochastic models (Beydoun et al. 2017; Broadley et al. 2012; Marcolli et al. 2007; Niemand et al. 2012; Vali and Stansbury 1966; Wheeler and Bertram 2012; Wright and Petters 2013). Stochastic models conceptualize the substance surface as having a stochastic distribution of active sites, each with specific characteristic catalytic strength. The models differ in the assumption of the functional form describing how the sites are distributed over the particle surfaces and in the parameters used to encode the catalytic strength of each active site. Contact angle, characteristic temperature, or free energy are often-used parameters. Active site models fit to experimental data can be convolved with surface area to predict INP (e.g., Niemand et al. 2012; Phillips et al. 2008).

Heterogeneous freezing nucleation on a single site is both repeatable and random. This is demonstrated in
experiments that track the observed nucleation temperature through multiple freeze–thaw cycles of the same surface or active site (Durant and Shaw 2005; Kaufmann et al. 2017; Peckhaus et al. 2016; Vali 2008, 2014; Vali and Stansbury 1966; Wright and Petters 2013; Wright et al. 2013). These experiments demonstrate that the same site consistently induces freezing within a few degrees Celsius interval; however, sometimes the freeze–thaw cycle can systematically modify the active site. Cooling cycle experiments can be used to determine the slope of the apparent nucleation rate of the active site (Wright and Petters 2013; Wright et al. 2013). The narrower the standard deviation of the freezing temperature in cooling cycles, the stronger the increase in nucleation rate with cooling. Thus, a narrow standard deviation, which has been reported in a diverse range of laboratory studies, also implies that there is a weak dependence on cooling rate (Wright et al. 2013). However, some case studies of clouds have suggested that the atmosphere comprises a large number of inefficient INPs that activate slowly over the time period of several hours (Fridlind et al. 2012; Westbrook and Illingworth 2013; Yang et al. 2013).

Pruppacher and Klett (2010) summarize the main requirements for an INP as follows: 1) The particle must be insoluble, to provide a solid surface on which the ice germ forms. The only reported exception is observed nucleation on alcohol monolayers (Gavish et al. 1992), which has not yet been shown to be important for ice nucleation in the atmosphere. 2) INPs tend to be found in the subset of the atmospheric aerosol that exists at larger diameters than the typical peak of the number size distribution. This assertion is based on correlation of atmospheric INPs with particles > 500 nm in diameter (DeMott et al. 2010; Georgii 1959) and on the size distribution of residuals of crystals formed on IN active particles (DeMott et al. 2010). 3) The surface should provide sites for hydrogen bonds to form. This is supported by studies of organic compounds that nucleate ice (Fukuta 1966) as well as molecular dynamic simulations of active sites (Kiselev et al. 2017). 4) Ice nucleating surfaces must have a good crystallographic match to ice. This is supported by the success of the intuition used in the initial search for active INP (Vonnegut 1947) and subsequent work confirming ice nucleating activity for substances with a good crystallographic match, experiments with amino acids (Gavish et al. 1992), and molecular dynamics simulations (Zielke et al. 2016). Last, 5) the surface must have active sites, as demonstrated above. From the survey above it should be evident that quantitative predictions of INP activity of a surface from chemical composition or other typically measured characteristics are not yet available.

c. Composition

Ambient measurements provide some constraints on the dominant particle types serving as INPs. One method that has been applied to isolate INPs from the total atmospheric aerosol population is to collect ice crystals through a counterflow virtual impactor (CVI; Noone et al. 1988), sublimate the ice, and analyze the residuals using single-particle mass spectrometry. This method is most accurate in the upper free troposphere below the homogenous freezing temperature of pure
water, where all cloud particles are frozen. In this regime, mineral dust and metallic particles dominate the INP. Particles containing sulfate or organics are fewer than in the clear-sky aerosol. Elemental carbon and biological particles are rarely observed (e.g., Cziczo et al. 2013). In the mixed-phase regime, complications in applying the CVI method arise, as separating ice crystals from supercooled cloud droplets via a virtual impactor is difficult. Therefore, many measurements of INP active in this temperature range have focused on analyzing particles that formed ice inside a continuous-flow diffusion cloud chamber, collecting them via impaction and probing composition and morphology of collected particles using electron microscopy (e.g., Boose et al. 2016a,b; Y. Chen et al. 1998; DeMott et al. 2003; Pratt et al. 2009; Prenni et al. 2009a,b; Rogers et al. 2001; Welti et al. 2018). These studies show that the lower-tropospheric INP population is a mixture of mineral dust, metal oxide particles, and carbonaceous particles of biological and anthropogenic origins. Recently, development of an ice counterflow virtual impactor has been reported, and this device has been used to sample ice crystals directly from mixed-phase clouds (Mertes et al. 2007). This technique has been used together with electron microscopy and single-particle mass spectrometry to identify the composition of ice residual particles for different clouds (e.g., Cozic et al. 2008; Mertes et al. 2007; Schmidt et al. 2017; Wurzingen et al. 2015). Besides mineral dust, these measurements have identified metal oxides, black carbon, lead-bearing particles, and carbonaceous particles including those of biological origin, in the ice residuals. As discussed by Wurzingen et al. (2015), both methods—collection of particles behind an ice nucleation detection instrument and collection of ice crystal residuals—are prone to artifacts. It remains difficult to fully quantify the link between composition and ice crystal formation from these types of studies. Another means to link composition and INP is by source sampling. Recent examples are characterizations of INP from biomass burning (Petters et al. 2009b; Sassen and Khvorostyanov 2008), sea spray (DeMott et al. 2016; Wilson et al. 2015), and glassy secondary organic aerosols (Ignatius et al. 2016; Knopf et al. 2018; Wang et al. 2012). Urban sources do not appear to contribute substantially to atmospheric INP concentrations, and pollution may lead to a degradation of ice nucleating ability (Braham and Spyers-Duran 1974; J. Chen et al. 2018). Plausible mechanisms are the digestion of active sites by pollutants or sulfuric acid (Sullivan et al. 2010a; Kanji et al. 2013). However, acidity and coatings do not always destroy active sites (Kanji et al. 2019; Salam et al. 2007; Sullivan et al. 2010b). In either case, coating of dust particles by sulfuric acid may also increase their effect as giant CCN particles on warm-rain formation (Levin et al. 1996; Wurzler et al. 2000). Robust methods linking these emissions and atmospheric processing to ice formation in clouds remain to be developed.

d. Instrumentation

Almost all ice nucleation instrumentation is custom built and is developed by individual research groups. Several difficulties exist in designing instrumentation for the detection and characterization of INPs. First, most techniques are sensitive to a specific mode of nucleation, for example, immersion mode or contact mode. Second, as shown in Fig. 11-6, the concentration range for immersion mode nuclei spans 10 orders of magnitude, a dynamic range that no single instrument can span. At the lower limit, concentrations are $10^{-3}$ m$^{-3}$ air, thus necessitating the equivalent sampling of 1000 m$^3$ of air in order to detect a single nucleus. Third, there is no universally accepted particle standard to calibrate ice nucleation instruments. Finally, any custom-built instrumentation is susceptible to measurement artifacts that are difficult to identify with a single technique. Consequently, a wide range of apparatuses with different measurement principles have been built, starting in the 1950s. A review of these techniques is provided elsewhere (Cziczo et al. 2017; DeMott et al. 2011).

The difficulties related to INP measurement were recognized early on and resulted in three intercomparison workshops that took place in the 1960s and 70s (Vali 1975). With the revival of the field in the early 2000s, many new instruments were built. The fourth intercomparison workshop (DeMott et al. 2011) was held in 2007 and drew participation from eight instruments with five different design principles. An interesting outcome of the workshop was the identification of discrepancies of several orders of magnitude in the measured ice active fraction of a dust sample at a single temperature. Since that workshop, many more instruments have been constructed, including a revival of the drop freezing assay technique (Vali and Stansbury 1966). The Fifth Ice Nucleation (FIN) Workshop was held in 2015, after various preceding smaller-scale intercomparison efforts (DeMott et al. 2017, 2018; Hiranuma et al. 2015). A dust intercomparison project (Hiranuma et al. 2015) and phase 2 of the FIN workshop (DeMott et al. 2018) had 17 and 20 participating instruments. The reported agreement between the different techniques was often within approximately one order of magnitude for laboratory mimics and ambient aerosol. Given the 10 orders of magnitude concentration range in Fig. 11-6, this finding implies reasonable agreement within a few degrees Celsius. However, some systematic differences have been noted between
techniques (Emersic et al. 2015; Hiranuma et al. 2015) and the community is continuing collaborative efforts to better understand the measurement diversity. Furthermore, it is not yet known to what precision and accuracy INPs need to be known to assess aerosol impacts on ice clouds.

e. Aerosol--INP--ice crystal closure

As described in Stith et al. (2019), the development of observing systems utilizing platforms ranging from ground-based systems, to balloons, to research aircraft, was instrumental in advancing the field of cloud physics. Similar to the development trajectory for aerosol instrumentation, early samplers relied on impaction techniques to capture cloud particles and postprocessed the samples for information on size, phase, shape, and composition. The development of in situ cloud probes greatly increased the amount and type of data available, and helped spur rapid advancements in the understanding of cloud and precipitation microphysics (chapter 5 in Wendisch and Brenguier 2013). Cloud probe data were relied on to develop improvements to the representation of cloud microphysical processes in models.

With advancements in measuring technologies, opportunities arose to reexamine existing datasets, particular with respect to complex ice microphysics and how these processes were tied to the atmospheric aerosol. McFarquhar et al. (2007) reported significant shattering artifacts in airborne optical cloud probes, thus leading to questions whether the large ice enhancement ratios (over the crystal concentrations expected for primary ice formation) observed in the early 1960–80s were correct. Korolev et al. (2011; Fig. 11-18) showed that the problem affects tips of cloud probes of different designs and that the number of fragments per ice particle can approach 1000. Fortunately, the problem can be mitigated by modifications to the tip design. Recent aerosol–ice crystal closure studies were performed during the Ice in Clouds Experiment–Layer Clouds (ICE-L) and Ice in Clouds Experiment–Tropical (ICE-T) experiments. During ICE-L both measurements and ice crystal number concentrations predicted from either dynamic cloud models or static INP parameterizations were on the order of 0.1–1 L$^{-1}$ for clouds colder than $-30^\circ$C (Eidhammer et al. 2010). In contrast, during ICE-T, ice

---

FIG. 11-18. High-speed video images of the trajectories of ice particles bouncing from (a) the arm tips of a cloud imaging probe, (b) an optical array probe, and (c) a forward scattering spectrometer probe inlet tube. Frames are from high-speed videos that were taken in ice sprays in a wind tunnel at an airspeed of 80 m s$^{-1}$. Red areas in (a) and (b) highlight the sample volumes of the probes. Particles unaffected by bouncing and shattering appear as horizontal lines. Particles bounced inside the inlet in (c) are not visible due to the lack of illumination. [Figure and caption adapted from Korolev et al. (2011).]
crystal number concentration exceeded 100 L\(^{-1}\) in clouds with tops ranging from \(-8^\circ\) to \(-15^\circ\)C (Lawson et al. 2015). Lawson et al. (2015) used cloud probes with modified tips. Thus, neither shattering nor primary ice nucleation can explain these observations. These two studies are consistent with the data from Hobbs et al. (1980; Fig. 11-19) and earlier studies, and point to the possibility of as-yet-undescribed ice multiplication processes. The observed ice enhancement is largest at warm temperatures and small at temperatures \(< -25^\circ\)C.

Many of the measurements in the 1960s and 1970s implied the coincidence of large drops (>250 \(\mu\)m) with large ice enhancement ratios, suggesting an ice multiplication mechanism involving the freezing and shattering of large drops (Pruppacher and Klett 2010). Such a process has recently been captured on high speed video (Wildeman et al. 2017) and quantified using levitated drops (Lauber et al. 2018). In general, insufficient understanding of secondary ice production is limiting the ability to quantify aerosol effects on ice crystal concentrations and subsequent cloud properties, although as described below, some attention has been focused on anthropogenic perturbations to these relationships and the implications for climate forcing. A recent review of secondary ice production and recommendations for future directions are summarized by Field et al. (2017).

5. Cloud physics

As noted by Pruppacher and Klett (2010), “the period of [rapid] progress [in cloud physics] since the beginning of the 1940’s has not been characterized by numerous conceptual breakthroughs, but rather by a series of progressively more refined quantitative theoretical and experimental studies of previously identified microphysical processes.” In their article “The Microphysics of Clouds” in Reports on Progress in Physics, Mason and Ludlam (1951) summarized the state of knowledge as gleaned from laboratory, field, and theoretical investigations up to the first decades of the twentieth century. The authors begin by noting that macrophysical processes that “determine the large scale features of cloud development” (the three-dimensional distributions of relative humidity, temperature, and winds) are of equal importance to the better-characterized microphysical processes—“condensation, droplet growth and coalescence, ice crystal formation”—that are the focus of their report. Their comments reveal that by 1951, the coupling between cloud dynamics and microphysical processes was recognized as a key factor in the development of precipitation, although many details of how these interactions occurred in various types of cloud systems were not yet well understood.

Referring to the summary by Hobbs (1991; Fig. 11-7) of notable decadal milestones in research on cloud microphysics and precipitation processes, we note that each decade featured work in theory/modeling, laboratory studies, and field studies. We have therefore organized this section in a similar way, although noting that the extremely large body of work that is represented in each of these categories cannot be done justice here and we have of necessity presented only selected highlights. In doing so, we have relied heavily on a number of excellent, periodic summaries of progress, beginning with the 1957 Meteorological Research Reviews (Landsberg et al. 1957), particularly the chapters by Weickmann (1957) and by aufm Kampe and Weickmann (1957) covering the period 1951–55. By this time, the community had made use of aircraft to collect data on aerosol, cloud, and precipitation particles throughout the troposphere, supplementing prior observations from the ground, mountaintops, and balloons, and had explored variability across cloud types and between marine and continental locations. It had become clear that cloud physics had to move beyond an understanding of the microphysics of individual drops and crystals to include cloud formation and cloud dynamics, and that considerable effort had to be devoted to additional observations and advances in theoretical understanding of the latter processes (aufm Kampe and
Weickmann 1957). The power of the emerging technologies of radar, satellite observations, and the computer in providing the data needed to develop the “macro-scale” aspects of cloud physics was also recognized early on (Squires 1967; Hidy 1967). Figure 11-20 shows the status of the emerging understanding of the “meteorological physics of precipitation development” as put forth by Braham (1968). In the mid-1960s, Squires (1967) identified the two leading basic problems in cloud physics as 1) understanding of the initiation and development of the ice phase and its role in the formation of precipitation, and 2) the details of turbulent transfer, alternately referred to as “mixing” or “entrainment” phenomena, in impacting buoyancy and the shape of cloud particle distributions. Interestingly, these same topics are repeatedly identified as outstanding questions in subsequent periodic commissioned reviews of advances in cloud physics, beginning in 1979 (Cotton 1979; Hallett 1983; Beard 1987; Rogers and DeMott 1991; Rasmussen 1995). That these phenomena have long been identified as key uncertainties demonstrates both the complexity of these processes and the attendant difficulties in designing targeted observational and laboratory experiments that fully capture those complexities.

a. Advances in theory of cloud particle growth and collision/coalescence

Kampe and Weickmann (1957) note the considerable progress, since the mid-1940s, that had been made in understanding the size distributions, number concentrations, and composition of atmospheric particles, including “giant” (supermicron) particles. Laboratory and field studies of the production of sea salt nuclei by bubble bursting had already been published. The modification of water drop diffusional growth due to the presence of solutes—including effects of solution nonideality (e.g., McDonald 1953)—had also been established, and laboratory experiments testing numerical formulas linking supersaturation, aerosol characteristics, and drop formation had generally confirmed theoretical expectations for droplet activation.

The growth of nucleated cloud droplets and ice crystals by diffusion was described by Mason and Ludlam (1951) in their seminal work, The Microphysics of Clouds, which was followed by their text in 1957 (Mason 1957) and subsequent updates. In this work they summarize the early body of work that sought to model the growth of a population of particles into cloud (or ice) droplets and from there to precipitation sized particles that would have sufficient terminal velocity to fall out of clouds. A conundrum tackled by the community was that predicted droplet condensational growth rates from the diffusional growth equation that Mason had developed were too slow to lead to precipitation-sized droplets, in comparison with the temporal and spatial scales of actual clouds. Further, condensational growth of particles and initial nucleated droplets, even of quite disparate initial sizes, was predicted to rapidly form narrow drop size distributions (Hidy 1967). While collision–coalescence was recognized as supporting faster growth of the droplets than diffusion of water vapor alone (Schumann 1940), the most efficient process for growing droplets relied on differential settling velocities for droplets of sufficiently different sizes, and it was not clear how nature produced drops of these different sizes. Mason and Ludlam reviewed additional hypotheses, which included effects of entrainment of dry air on broadening the drop size distribution (e.g., Howell 1949)—namely, the “mixing” processes that have continued to received attention through the decades. Turbulent fluctuations were also recognized as promoting coalescence by increasing the chance of droplet collisions and hence increasing the collection kernel (e.g., Franklin et al. 2005; Ayala et al. 2008; Grabowski and Wang 2013). An additional early speculation (Mason and Ludlam 1951) was the probable existence of large (ultragiant) nuclei that could trigger collision–coalescence even if present in very low number concentrations that were not readily observable at the time. Subsequently, drizzle formation was found to be enhanced in the presence of giant CCN, such as sea salt or mineral dust, in observational studies (Rosenfeld et al. 2002; Hudson et al. 2011); their impacts were confirmed in modeling studies by, for example, Feingold et al. (1999), Saleeby and Cotton (2004), and Posselt and Lohmann (2008).

An important advance in the conceptual understanding of warm-rain processes was presented by Berry and Reinhardt (1974a,b,c,d). Observations had shown that cloud droplets must surpass ~20 μm in radius before collision–collection became rapid enough to grow drops to precipitation sizes on typical cloud time scales. Twomey (1964) presented the first numerical integration of the statistical collection equations for a continuous size distribution, showing that small numbers of large drops were produced rapidly. By undertaking a more detailed computational solution of the stochastic collection equation in combination with condensational growth, Berry and Reinhardt showed the conditions under which a large-droplet mode could be generated within ~15 min, providing the needed separation in drop size space to initiate rapid collection and leading to drizzle production on realistic time scales. Throughout and after these studies, attention also focused on laboratory experiments (as summarized in Weickmann 1957, and in many later works as described briefly below) to
Fig. 11.20. (top) Conceptualization of precipitation processes. (bottom) Interactions and feedback loops between precipitation physics and cloud dynamics. [Figures from Braham (1968).]
determine the appropriate values of various collection kernels needed in the theoretical descriptions. Uncertainties in growth by collision–coalescence are related to the collection efficiency, which is the product of the collision efficiency times the coalescence efficiency. For collisions involving drop radii below ~50 μm, assuming a coalescence efficiency of 1 is a good assumption (e.g., Beard et al. 2002). The gravitational collision efficiency depends on the sizes of the collector drop and the droplets that are collected, such that it increases for larger collector drops and for an increasing ratio of the smaller to the larger drop (e.g., Schlamp et al. 1976).

At midlatitudes, observations suggested that heavier precipitation often formed only after the initiation of an ice phase, generally from supercooled liquid drops. Crystal habits were found to be correlated with atmospheric temperatures and supersaturations (Nagaya 1954; Magono and Lee 1966). Theoretical explanations of how the various complex habits of ice crystals grew in the atmosphere were put forward, based on the possible crystalline structures of ice and theories of growth of organized crystal faces or at surface dislocations, supported by numerous early laboratory studies. An explanation for how ice formation could lead rapidly to the formation of precipitation-sized droplets was put forward well before 1950 (Weickmann 1957). It was recognized that after the first ice crystals have been formed, the cloud becomes thermodynamically unstable because of the lower water vapor pressure over ice than over liquid water. If the water vapor pressure lies between saturation with respect to liquid water and ice, then the ice crystals will grow rapidly at the expense of the cloud droplets. This process is called the Wegener–Bergeron–Findeisen (WBF) process (Wegener 1911; Bergeron 1935; Findeisen 1938) and is of vital importance for weather and climate (List 2006; Storelvmo et al. 2015).

In the tropics, precipitation accounts for 69% of the total precipitation (Lau and Wu 2003). In midlatitudes, warm rain is even less prevalent, especially over land. Here it accounts for less than 10% of the total precipitation because of the smaller cloud droplets in continental clouds that are less likely to grow by collision–coalescence as compared to the larger cloud droplets in marine clouds (Mühlmenstädt et al. 2015).

Marshall and Palmer (1948) published results of measurements of raindrop distributions made by collecting drops onto dyed filter papers, and suggested that these could be fit by an exponential distribution (Fig. 11-21). The computational simplicity of the Marshall–Palmer distribution has led to its widespread application in models. Further, the relationship was found to be consistent with theoretical predictions for steady-state raindrop distributions and applicable to large sample sizes; however, it is likely not valid for short periods and especially at low rain rates (Cotton 1979; Hallett 1983; Rasmussen 1995).

b. Advances in laboratory studies

Numerous early (prior to 1950) laboratory studies are summarized in Mason and Ludlam (1951), primarily small-scale diffusion or expansion chambers combined with microscopy techniques that examined drop and crystal formation and growth and counted numbers of drops or ice crystals nucleated on samples of atmospheric aerosols; additional work during the 1950s and 1960s, that attempted to go beyond single-particle experiments, was also summarized by Hidy (1967). Interest in linking aerosols to warm and cold cloud formation continued into subsequent decades, with particular emphasis on studying population dynamics in large expansion chambers. The Colorado State University (CSU) Dynamic Cloud Chamber (DeMott et al. 1983) and University of Missouri at Rolla expansion chambers (White et al. 1987) were used to study cloud drop and ice formation under temperature, humidity, and simulated updraft conditions representative of cloud types ranging from fogs to cirrus. Figures 11-22

FIG. 11-21. Distribution function proposed by Marshall and Palmer (solid straight lines) as fits to the data from their Ottawa observations (dotted lines) and earlier data from Laws and Parsons (broken lines). The solid lines are fit to the functional form of an exponential distribution $N_D = N_0 \exp(-\lambda D)$, where $N_D$ is the value of the drop distribution function, $D$ is the drop diameter, $N_0$ is the value of $N_D$ at $D = 0$ and was proposed as 0.08 cm$^{-2}$ for any intensity of rainfall, and $\lambda$ was related to the rainfall rate $R$ (mm h$^{-1}$) by the following equation: $\lambda = 41 R^{-0.21}$ cm$^{-1}$. [Figure and caption adapted from Marshall and Palmer (1948).]
and 11-23 show the characteristics of these facilities, along with others described briefly below.

Recognizing the limitations of the classes of chambers described above to adequately address particle–particle interactions, especially for large drops that could not remain suspended for a sufficient time, a variety of wind tunnels were developed to study hail growth and droplet collision, coalescence, and deformation. In the United States, the University of California, Los Angeles (UCLA), Cloud Physics Laboratory, and particularly the UCLA Cloud Tunnel, which was established by Hans Pruppacher beginning in the mid-1960s, was one of the first to enable studies of cloud physics wherein hydrometeors were allowed to interact in a more realistic way with moving air, under controlled temperature and humidity conditions (Flossmann et al. 2010). Key findings from this facility included measurements of the terminal velocities and drag coefficients of small water drops, raindrop fall speeds and deformation behaviors, and growth and hydrodynamics of water and ice particles. The facility, and a later version at the University of Mainz, also contributed data on the role of turbulence and on the role of clouds in atmospheric chemistry, through particle and gas scavenging studies. Kenneth Beard, who had studied at UCLA under Pruppacher, helped found the cloud physics laboratory at the Illinois State Water Survey (Changnon and Huff 1997). Notable contributions from this laboratory included measurements of cloud drop collision and coalescence (e.g., Beard and Ochs 1983), providing critical data on the kernels needed to model these processes. Information on raindrop shape derived from laboratory studies and numerical computations has been key to improvements in the use of radar to estimate rain rates (e.g., Beard et al. 2010). Roland List, working first at the Swiss Federal Institute for Snow and Avalanche Research and later at the University of Toronto, developed a series of wind tunnels that were used to study hail formation.
<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Reference</th>
<th>Location</th>
<th>Operational since</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerosol Interaction and Dynamics in the Atmosphere (AIDA) chamber</td>
<td>Expansion-type cloud chamber</td>
<td>Möhler et al. (2003)</td>
<td>Karlsruhe, Germany</td>
<td>1996–present</td>
<td>Aerosol and cloud chemistry, cloud microphysics (especially ice nucleation studies), aerosol and cloud radiative properties</td>
</tr>
<tr>
<td>Calspan chamber</td>
<td>Expansion-type cloud chamber</td>
<td>Hoppel et al. (1994)</td>
<td>Ashford, New York</td>
<td>1980s–present</td>
<td>Aerosol and cloud chemistry, cloud microphysics, aerosol and cloud radiative properties</td>
</tr>
<tr>
<td>Experimental Multiphasic Atmospheric Simulation Chamber (CESAM)</td>
<td>Multiphase reaction chamber</td>
<td>Wang et al. (2011)</td>
<td>Créteil, France</td>
<td>~2009–present</td>
<td>Aerosol and cloud photochemistry</td>
</tr>
<tr>
<td>Colorado State University (CSU) chamber</td>
<td>Adiabatic expansion-type cloud chamber</td>
<td>DeMott and Rogers (1990)</td>
<td>Fort Collins, Colorado</td>
<td>Out of operation</td>
<td>Cloud microphysics (especially ice nucleation studies)</td>
</tr>
<tr>
<td>Cosmics Leaving Outdoor Droplets (CLOUD) chamber</td>
<td>Multiphase reaction chamber</td>
<td>Duplissy et al. (2010)</td>
<td>European Organization for Nuclear Research (CERN), France/Switzerland</td>
<td>2006–present</td>
<td>Influence of galactic cosmic rays on particle formation and aerosol chemistry, cloud chemistry and microphysics</td>
</tr>
<tr>
<td>Desert Research Institute (DRI) chamber</td>
<td>Adiabatic expansion-type cloud chamber</td>
<td>Stehle et al. (1981)</td>
<td>Reno, Nevada</td>
<td>Out of operation</td>
<td>Cloud chemistry</td>
</tr>
<tr>
<td>Energy Research Centre of the Netherlands (ECN) high-flow cloud chamber</td>
<td>Turbulent cloud wind tunnel</td>
<td>Khlystou et al. (1996)</td>
<td>Petten, the Netherlands</td>
<td>1996–present</td>
<td>Aerosol–cloud interactions</td>
</tr>
<tr>
<td>Manchester Ice Cloud Chamber (MICC)</td>
<td>Fall tube</td>
<td>Connolly et al. (2012)</td>
<td>Manchester, United Kingdom</td>
<td>2009–present</td>
<td>Cloud microphysics (e.g., ice nucleation and aggregation), cloud radiative properties, and thunderstorm electrification</td>
</tr>
<tr>
<td>Meteorological Research Institute (MRI) chamber</td>
<td>Adiabatic expansion-type cloud chamber</td>
<td>Tajiri et al. (2013)</td>
<td>Tsukuba, Japan</td>
<td>2012–present</td>
<td>Cloud microphysics (cloud droplet/ice crystal formation and growth), aerosol scavenging</td>
</tr>
<tr>
<td>Penn State University chamber</td>
<td>Mixing chamber</td>
<td>Song and Lamb (1994)</td>
<td>University Park, Pennsylvania</td>
<td>Out of operation</td>
<td>Cloud chemistry and microphysics</td>
</tr>
<tr>
<td>University of Manchester Institute of Science and Technology (UMIST) chamber</td>
<td>Fall tube</td>
<td>Latham and Reed (1977)</td>
<td>Manchester, United Kingdom</td>
<td>Out of operation</td>
<td>Cloud mixing processes</td>
</tr>
<tr>
<td>University of Missouri–Rolla (UMR) chamber</td>
<td>Adiabatic expansion-type cloud chamber</td>
<td>Hagen et al. (1989)</td>
<td>Rolla, Missouri</td>
<td>Out of operation</td>
<td>Cloud microphysics</td>
</tr>
</tbody>
</table>

Fig. 11.23. A representative list of cloud chambers worldwide focused on cloud and aerosol physics research. [Figure from Chang et al. (2016).]
(e.g., List 1966) and drop breakup behavior (McTaggart-Cowan and List 1975), the latter process recognized as an important mechanism shaping drop size distributions and influencing precipitation formation (Blanchard 1950). Later advances in laboratory studies of these phenomena were discussed by Beard (1987) and Rogers and DeMott (1991).

As noted in Fig. 11-7, during the 1950s considerable effort was directed toward advancing understanding of all topics in cloud physics, largely motivated by the possibility of weather modification (Squires 1967) by artificially inducing cloud glaciation, a possibility raised by the discovery of the WBF process and the work by Schaef er, Vonnegut, and Fletcher (Fletcher 1961; Schaefer 1946; Vonnegut 1947) on INPs. The National Academy of Science report “Weather and Climate Modification” (National Research Council 1973) also spurred increased investment in research and applications aimed at these topics. Thus, laboratory ice nucleation research in the 1940–80 period was strongly motivated by weather modification applications, and sought to understand the abundance and sources of natural INPs, to identify and produce suitable agents to nucleate ice, and to understand the link between INPs and cloud ice crystal number concentration. As summarized by Mossop (1970), the growing cloud physics community had realized early on (e.g., Dobson 1949) that ice particles were found in the atmosphere at warmer temperatures and in higher concentrations than could be reconciled with laboratory measurements of INPs; by the late 1960s, evidence of a substantial mismatch between ice crystal number concentration and INP concentrations of up to five orders of magnitude had emerged (Auer et al. 1969; Koenig 1963; Mossop 1970; Hobbs et al. 1980, Fig. 11-19). Mechanisms for ice initiation and multiplication were put forward and tested in laboratory settings. Secondary ice production by rime splintering was discovered by Hallett and Mossop (1974); their work stands as a major contribution in the last half century to modern understanding of cloud physics. Cotton (1979), in his review for the 1975–79 International Union of Geodesy and Geophysics (IUGG) Quadrennial Report, noted that “one of the most significant advances in cloud physics during the past four years was the confirmation and refinement of the Hallett–Mossop ice multiplication mechanism,” work done in Mossop’s laboratory at the CSIRO in Australia. Hallett also continued to produce seminal work on ice microphysics after founding the Ice Physics Laboratory at the Desert Research Institute in the late 1960s (Fig. 11-22).

The discovery of the importance of secondary ice production in many natural cloud systems led to a hiatus in active ice nucleation research (Fig. 11-3) due to the perceived irrelevance of INPs in controlling cloud properties and precipitation initiation. For decades, the connection between the atmospheric aerosol and ice initiation in modestly supercooled clouds was lost. Concerns about the role of anthropogenic aerosols in climate change via glaciation indirect effects (Lohmann 2002; Lohmann and Feichter 2005) and the discovery that some airborne ice crystal concentration measurements may have been significantly overestimated by shattering artifacts (McFarquhar et al. 2007) led to renewed and ongoing interest in ice nucleation research (DeMott et al. 2011). In addition, surface sources of ice crystals such as blowing snow and hoar frost were recently identified as contributors to ice crystal number concentrations at mountain top measurement sites (Lloyd et al. 2015; Farrington et al. 2016; Beck et al. 2018). Secondary ice formation processes have also received renewed attention, as summarized in the recent review by Field et al. (2017). In addition to the Hallett–Mossop process, splinter production following the freezing of large droplets (Leisner et al. 2014; Lauber et al. 2018) and collision-induced fragmentation (Knight 2012) have been identified through laboratory studies as possible secondary ice formation mechanisms.

By 1985, sufficient concern around the lack of investment in experimental cloud physics and cloud chemistry facilities in the United States, particularly as compared with support for modeling and observational studies, had led to the organization of a technical workshop to review current capabilities and recommend future directions (List et al. 1986). The workshop report stressed that “laboratory research in cloud physics and cloud chemistry has been a very productive and necessary complement to field and modeling studies,” with strong potential for continuing fundamental contributions in the atmospheric and related sciences. An identified limitation was that no existing facilities (see compilation in Fig. 11-22) could accommodate study of the interactions of more than one particle type (e.g., drops and ice crystals) nor interactions with gases for the purposes of studying cloud chemistry; the report called for investments in new capabilities and consideration of establishment of a national facility (see also Beard 1987). Most recently, Chang et al. (2016) reported development of a new type of cloud chamber that can achieve steady-state conditions in the laboratory and that can be used to address the more complex interactions identified by List et al. (1986) as outstanding unresolved processes. Chang et al. (2016) also presented a compilation of laboratory capabilities similar to that composed by List et al. (1986), shown in Fig. 11-23. Interestingly, a number of the facilities
mentioned in the 1986 report are shown as out of operation by 2016. They have been replaced to a certain extent by advanced facilities outside of the United States that have hosted larger, international efforts that are more suited to current outstanding needs in cloud physics. Notable examples include the intercomparisons of INP measurement techniques that have been undertaken at the Aerosol Interaction and Dynamics in the Atmosphere (AIDA) chamber in Karlsruhe, Germany (DeMott et al. 2018); the development of the Leipzig Aerosol Cloud Interactions Simulator (LACIS; Stratmann et al. 2004), designed to probe aerosol–cloud interactions in the difficult-to-access region very near water saturation; and the Cosmics Leaving Outdoor Droplets (CLOUD) experiment at the European Organization for Nuclear Research [Conseil Européen pour la Recherche Nucléaire (CERN)] that examines aerosol formation and impacts on clouds (Dunne et al. 2016).

c. Advances in observational studies

As noted in previous chapters of this monograph, it was recognized very early in the development of cloud physics as a discipline that advances in understanding cloud properties and precipitation development hinged on the strong coupling between cloud physics and dynamics. Hence, observational studies were needed to make progress. The chapter by Stith et al. (2019) outlines progress in the development and application of atmospheric observing systems. It is well beyond the scope of this review to summarize the numerous field studies that have been undertaken that included measurements of aerosols, cloud particles, and precipitation that have led to new insights. Here we mention a few exemplary studies of different cloud system types, and note that other chapters in this series cover many of these in greater depth.

The GARP Atlantic Tropical Experiment (GATE; https://www.ametsoc.org/sloan/gate/; https://www.eol.ucar.edu/field_projects/gate) was the first major experiment of the Global Atmospheric Research Program, an early (1960s–70s) international effort (72 nations) to organize large-scale, coordinated studies to address outstanding problems in atmospheric science. The focus of GATE was largely on improving weather predictions, and thus of necessity included radar observations and cloud physics measurements on aircraft. The observational dataset provided an unprecedented look at cloud and atmospheric structure (e.g., as summarized in Hallett 1983) and has been used in numerous modeling studies. The Hawaiian Rainband Project (HaRP) in 1990 (https://www.eol.ucar.edu/field_projects/harp1990) was aimed at studying the factors that led to heavy precipitation in Hawaiian convection in the absence of an ice phase, specifically targeting the interactions between cloud dynamics and microphysics. In contrast, the Winter Icing in Storms Project (WISP; Rasmussen et al. 1992; Rasmussen 1995) was conducted along the Colorado Front Range in the winters and early springs of 1990 and 1991, and sought to elucidate the dynamical and microphysical processes governing the budget of supercooled liquid water in winter storms.

By the time of the review by Hallett (1983), the radiative properties of clouds were attracting renewed interest, due to their impacts on climate. The work of Twomey (1974, 1977) had suggested the link between CCN, cloud droplet number concentrations, cloud drop effective radius, and cloud albedo; ship tracks that are sometimes visible as bright lines behind ships detected in early satellite imagery appeared to be direct manifestations of these relationships (Coakley et al. 1987; Russell et al. 1999), and air pollution in the form of increased atmospheric aerosol loadings was postulated to have global radiative impacts through modifying cloud properties on larger scales. The First ISCCP Regional Experiment (FIRE; Cox et al. 1987; Rasmussen 1995) was organized to study the role that clouds play in the global climate, focusing on marine stratocumulus and cirrus. Particularly for these cloud types, radiative properties are strongly linked to microphysics, and hence microphysical observations were an important part of the measurement strategy. Among many other important findings, observations of two ship tracks during FIRE showed smaller droplet sizes and higher liquid water contents than in background stratocumulus cloud, suggesting anthropogenic perturbations of cloud radiative properties and hence an impact on climate. Factors affecting drizzle initiation in stratocumulus clouds continued to be a focus of observational studies for several decades, due to their postulated microphysical sensitivity to perturbations and thus their central role in anthropogenic indirect effects on climate. The Eastern Pacific Emitted Aerosol Cloud Experiment (E-PEACE;Russell et al. 2013) deliberately modified stratocumulus clouds with two different particle types to investigate the role of particle composition and to separate dynamical effects of ship exhaust from microphysical changes due to the aerosol source strength.

Stephens et al. (2019) discuss how the advent of satellite-borne instruments has enabled the diagnosis of cloud microphysical properties from space, providing near-global coverage of many different cloud types as well as decades of data from which trends can be deduced (Stith et al. 2019; Ackerman et al. 2019). Integrated observational strategies (e.g., Stephens et al. 2018) permit not only snapshots of cloud states, but can
be used to derive information on processes themselves (e.g., Müllmenstäd et al. 2015). These data are invaluable for diagnosing reasons for model biases, for example, the too-frequent light rain bias in many global models (Stephens et al. 2010). Figure 11-24, adapted from Suzuki et al. (2015), demonstrates this concept by relating the gradient in radar reflectivity (from the CloudSat spaceborne radar) to the gradient in cloud optical depth; a zero gradient implies no droplet collection and hence no drizzle. Figure 11-24 reveals that the model tends to drizzle even at low droplet sizes, whereas the satellite observations show that drizzle onset does not occur until effective radii exceed 10 μm.

**d. Advances in numerical modeling of cloud microphysics**

The first cloud microphysical parameterization was put forward by Kessler (1969) and tested in a two-dimensional model. At that time, cloud fraction and cloud optical depth were often prescribed as a function of latitude (Manabe et al. 1965; Boer et al. 1984). More advanced methods diagnosed cloud cover of stratiform clouds from grid-mean values of relative humidity (Smagorinsky 1960; Geleyn 1981) or by assuming a probability distribution function of the total water (e.g., Smith 1990; Sommeria and Deardorff 1977). Approaches also considered the inversion strength for marine boundary clouds forming in regions with persistent subtropical highs (Slingo 1987; Klein and Hartmann 1993). Such approaches appeared in GCMs in the 1980s and 1990s (e.g., Slingo 1987; McFarlane et al. 1992; Kiehl et al. 1994). More advanced methods include cloud parameterizations that obtain cloud water/ice from the respective transport equation following the pioneering work of Sundqvist (1978). In some global climate models, cloud cover remains parameterized mainly as a function of relative humidity (e.g., Stevens and Bony 2013) as more advanced cloud cover schemes such as Tompkins (2002) produce an unrealistically high climate sensitivity. In current models from cloud scale to global scale, cloud microphysical processes are parameterized at very different levels of complexity. Khain et al. (2015) discuss the historical development and application of bulk microphysical parameterizations and bin microphysics approaches, applied to models at these various scales. As detailed in their compilation, bulk approaches range from solving prognostic equations for not only cloud liquid water and cloud ice but also for
rain, snow, graupel, and hail (e.g., Fowler et al. 1996; Gettelman and Morrison 2015); to using two-moment schemes especially in order to study aerosol–cloud interactions (e.g., Lohmann 2002; Gettelman et al. 2008; Salzmann et al. 2010); to using four moments but only a single category ice class (e.g., Morrison and Milbrandt 2015; Dietlicher et al. 2019). “Bin” representations (e.g., Morrison and Grabowski 2007) and “superdroplet” approaches (e.g., Shima et al. 2009) explicitly resolve hydrometeor category evolution via microphysical processes. Cotton (2003) presented an overview of cloud model development through 2000 that included a discussion of the “bin-emulating” microphysics package developed for the CSU Regional Atmospheric Modeling System (RAMS), which has been further developed to address microphysical details of aerosol–cloud interactions while remaining computationally tractable for high-resolution modeling (Saleeby and van den Heever 2013). Grabowski et al. (2019) present an overview of the development of these various numerical schemes applied to the representation of cloud microphysics in detailed, cloud-scale models, including a discussion of the relevant merits and shortfalls of the various approaches.

e. Advances in understanding of aerosol influence on cloud systems

Increasing computational power enabling high-resolution atmospheric simulations, increasingly complex modeling frameworks that permit detailed representations of aerosol–cloud microphysical processes, and data from observational studies aimed at improving understanding of both cloud dynamics and cloud microphysics, have combined to enable significant advances in understanding of how the microphysical phenomena described in the preceding sections interact with atmospheric dynamics and thermodynamics in various cloud systems. The resulting advances in process-level understanding of aerosol–cloud system interactions have then been synthesized for representation in large-scale climate models, in order to estimate global impacts of aerosols on planetary albedo and precipitation, as discussed in the next section. Fan et al. (2016) review studies of aerosol–cloud interaction mechanisms, and additional compilations for various individual cloud systems have also been published. In this section, we make use of these reviews to briefly describe the current state of understanding of aerosol–cloud interactions in the most-studied cloud systems, where cloud microphysical processes have been treated along with dynamical feedbacks and in the context of specific environmental conditions that also affect how cloud systems evolve and respond (Gettelman and Sherwood 2016).

f. Marine stratocumulus

The most extensively studied cloud regimes are low-level warm clouds, and in particular, marine stratocumulus clouds (Wood 2012), and hence understanding of aerosol–cloud interactions in these systems is more advanced than for other cloud types (Fan et al. 2016). [The special case of feedbacks in trade wind shallow cumulus clouds is discussed in Vial et al. (2017)]. Stratocumulus clouds are driven by buoyancy induced by cloud-top radiative cooling and are characterized by a temperature inversion at cloud top that constrains their vertical development (Fig. 11-25). Their properties and evolution are further determined by surface fluxes and by microphysical characteristics, especially whether or not drizzle forms (Fig. 11-26). Early interest in stratus and stratocumulus arose from the postulated aerosol-mediated radiative and microphysical feedbacks on these clouds put forth by Twomey (1974), and others, as well as the large role that persistent stratus and stratocumulus decks play in Earth’s radiation budget, which enhances the importance of even small changes in reflectivity and persistence of these clouds. A cloud lifetime effect (now known as the second aerosol indirect

FIG. 11-25. Schematic showing the key processes occurring in the stratocumulus-topped boundary layer. [Figure and caption from Wood (2012).]
effect) was put forward by Albrecht (1989) and was based on the hypothesis that there are two stable regimes in the boundary layer (Baker and Charlson 1990): one in which the cloud consists of few but large cloud droplets that readily grow by collision–coalescence to drizzle size and thereby remove the aerosols that acted as CCN. This process keeps the boundary layer clean and the aerosol concentration low. In the other regime, the boundary layer is characterized by a high aerosol concentration, so that the cloud droplets are more numerous and smaller. These clouds do not readily drizzle and the aerosol concentration in the boundary layer can thus further accumulate.

Stevens et al. (2005) report the discovery of pockets of open cells (POCs), regions of open cellular convection embedded in otherwise uniform stratocumulus. The POCs correspond to long-lived features with increased drizzle and reduced cloudiness. Observations show that CCN concentrations are strongly reduced inside POCs, leading to the hypothesis that aerosols are important contributors governing the formation and maintenance of POCs (Petters et al. 2006; Sharon et al. 2006; Wood et al. 2008). Goren and Rosenfeld (2012) suggested that anthropogenic emissions can cause a change from the clean regime that is typically associated with open cells in stratocumulus decks to a closed-cell regime without...
drizzle and with more polluted conditions in the boundary layer. Open- and closed-cell regimes have been observed and analyzed (e.g., Feingold et al. 2015), confirming the role of precipitation and aerosol–microphysical feedbacks in transitions between the two regimes. However, small-scale studies and in situ observations do not show a generalized, systematic increase in cloud lifetime as concentrations of CCN increase, because of enhanced entrainment in the polluted clouds and because of faster evaporation of the smaller droplets (Jiang et al. 2006; Bretherton et al. 2007; Small et al. 2009). Entrainment and evaporation also affect the local environment, creating additional feedbacks (e.g., Lee et al. 2014; Seifert and Heus 2013). Further, the availability of even low number concentrations of giant CCN (GCCN) has been shown in modeling studies to transform nonprecipitating stratocumulus systems into precipitating systems (Feingold et al. 1999; Laird et al. 2000). At the global scale, and consistent with the above findings, Chen et al. (2014) conducted an analysis of satellite data for warm clouds and determined that tropospheric stability and humidity in the free troposphere, along with whether or not the clouds were precipitating, controlled the response of cloud liquid water contents to increases in aerosol loadings. However, synoptic-scale dynamics also play a role in modulating cloudiness. For example, Yuter et al. (2018) showed that gravity waves can rapidly clear large areas of stratocumulus in the subtropical southeast Atlantic.

Ship tracks, discussed above, have been investigated as a means to isolate a strong, localized aerosol–cloud interaction to test hypotheses for mechanisms that may be operative more broadly in low-level warm clouds. Shown in Fig. 11-27 are data for a “classical” ship track that was observed in the Monterey Area Ship Track (MAST) study (Durkee et al. 2000a), where the tenfold increase in aerosol number concentration led to an increase in cloud droplet number concentration of a factor of 3, a decrease in effective cloud droplet radius by 2 μm and an increase in cloud albedo from below 0.5 to more

![Figure 11-27](image_url)
than 0.6 (Durkee et al. 2000b). However, the global ship track coverage only amounts to 0.002% (Schreier et al. 2007), because formation of persistent ship tracks requires a specific structure of the boundary layer, in which clean background CCN concentrations are present, the cloud layer is thermodynamically and dynamically coupled to the surface, little wind shear is present, and a moisture inversion within specific ranges exists. The advance of satellite observations enabled a global view of the importance of ship tracks. Schreier et al. (2007) analyzed one year of AATSR data. They found a high temporal variability in ship track occurrence such that the global annual mean radiative forcing from ship tracks amounts only to a range from $2^{0.4}$ to $2^{0.6}$ mW m$^{-2}$, which is negligible compared to model estimates of ship tracks discussed below. Similar conclusions were reached from a 3-yr analysis of MODIS satellite data (Peters et al. 2011) such that no statistically significant impacts of shipping emissions on large-scale cloud fields could be identified. Christensen and Stephens (2011) offered an explanation for this from 2.5 years of CALIOP space-borne lidar data of ship tracks off the west coast of North America (20°–60°N and 150°–110°W). While these observations confirmed the dominant role of microphysical effects from ship tracks increasing cloud liquid water and cloud optical depth in the open-cell regime, as found, for example, by Durkee et al. (2000b), they found the opposite effect with decreases in cloud liquid water in response to ship exhaust in the closed-cell regime.

g. Arctic mixed-phase clouds

Mixed-phase clouds are of special interest not only because of their key roles in Arctic climate feedbacks, but also because of their surprising longevity that is not consistent with simplified views of cloud phase transitions (Morrison et al. 2012; Fig. 11-28). Although Arctic mixed-phase clouds share some structural characteristics with warm marine stratus clouds in that radiative cooling drives buoyancy and large-scale subsidence plays a role in their maintenance, mixed-phase clouds are inherently thermodynamically unstable due to the difference in water vapor pressure over liquid water and ice. If the ambient water vapor pressure is in between saturation with respect to water and ice, then the Wegener–Bergeron–Findeisen mechanism is expected to prevail, leading to complete glaciation of the cloud (Korolev 2007). Thus, for a mixed-phase cloud to be long-lived, either a constant source of cloud droplets is needed (e.g., Lohmann et al. 2016b), or the ice crystals need to leave the cloud faster than the time required for complete glaciation. The latter is the case in Arctic mixed-phase clouds. Their longevity arises because the few ice crystals that nucleate at cloud top due to longwave cooling,
grow rapidly in the water-saturated environment and sediment out of the cloud, leaving a layer of supercooled cloud droplets at cloud top behind (Fig. 11-28).

The cloud radiative properties that drive the dynamics are in turn sensitive to the relative amounts of liquid and cloud water, and thus these systems respond strongly to changes in their microphysical properties. Model simulations of Arctic mixed-phase clouds and their responses to aerosol perturbations reveal that even large-eddy simulations (LES) that simulate the dynamics correctly have problems in simulating observed INP and ice crystal number concentrations (e.g., Possner et al. 2017). Despite this, the study showed that increasing CCN induced a stronger cloud-top cooling that favored ice formation. Solomon et al. (2018) studied the sensitivity of Arctic mixed-phase clouds to the availability of both CCN and INP, the latter required for conversion of supercooled droplets to ice since temperatures are above the homogeneous freezing threshold. They found that treating both aerosol types as prognostic variables results in vertical sorting of the aerosols, with increased concentrations of CCN above cloud top that serve as a source of liquid drop formation, and increased concentrations of INPs at cloud base, preventing rapid glaciation as noted above. However, further work on simulating these cloud types is needed: an intercomparison study of the 2008 Arctic Summer Cloud Ocean Study (ASCOS) campaign involving LES and numerical weather prediction models reveals huge model-to-model differences in the response of Arctic mixed-phase clouds to increases in aerosol concentrations, because of differences in the representation of the cloud droplet size distribution among models (Stevens et al. 2018).

Interestingly, observations of ship tracks in Arctic mixed-phase clouds reveal a much smaller aerosol impact on mixed-phase clouds than on warm clouds (Christensen et al. 2014). While in warm clouds the precipitation formation rate via the warm phase is substantially reduced due to anthropogenic aerosols, precipitation formation via the ice phase is enhanced in mixed-phase clouds due to ship pollution decreasing the total cloud water in the tracks. This decrease competes with the decrease in effective radius such that the increase in cloud optical depth is much smaller than in warm clouds where the cloud water remains virtually unchanged.

**h. Orographic clouds**

There have been some systematic studies to evaluate the differences in orographic precipitation between clean and polluted conditions (e.g., Jirak and Cotton 2006; Rosenfeld et al. 2007; Muhlbauer and Lohmann 2009; Zubler et al. 2011). Those studies generally found that a decrease in warm-rain formation in orographic clouds causes some precipitation to spill over to the leeward side of the mountain, while cold-rain formation is less affected. Decreasing snowfall rates with increasing anthropogenic aerosol loads have been observed in the Rocky Mountains (Borys et al. 2000, 2003). The authors suggested that this decrease in snowfall is caused by a reduction in the collision efficiency of snowflakes with the smaller cloud droplets, which in turn reduced the riming rate. A reduced riming rate has also been found in model simulations to lead to spillover of some orographic precipitation to the leeward side of the mountain (Zubler et al. 2011). However, in another model simulation the more efficient WBF process under polluted conditions was found to compensate for most of the reduction in riming, causing only a minimal net change in precipitation (Saleeby et al. 2013). Observational studies in the Sierra Nevada also found that variability in both anthropogenic and natural aerosols (including variability in INPs) could impact cloud phase as well as the amount and location of precipitation (Rosenfeld et al. 2008b; Creamean et al. 2015).

The above studies, along with numerous other modeling and observational studies of aerosol impacts on orographic clouds, have been summarized in a recent review by Choudhury et al. (2019). They provide a detailed discussion of the current state of understanding that leads to the proposed suite of aerosol–orography–precipitation (AOP) interactions that is shown schematically in Fig. 11-29, resulting in precipitation changes on both the windward and leeward sides of the mountain. These authors call for increased attention to this topic to test the underlying microphysical interaction hypotheses, as much of the world’s supply of freshwater arises from orographic precipitation. If anthropogenic aerosols impact the microphysical evolution of clouds, and in turn shift the distribution of orographic precipitation in certain regions, and especially if they alter the watersheds into which this precipitation falls, the resulting impacts on agriculture and human health could be extreme.

i. Cirrus clouds

A comprehensive review of the macrophysical and microphysical properties of cirrus clouds is provided by Heymsfield et al. (2017), and a review of the response of these high-level, ice-phase clouds to anthropogenic perturbations is provided by Kärcher (2017). Anvil cirrus comprise ice crystals detrained from deep convective (see section below) or frontal clouds; cirrus clouds can also arise from homogeneous or heterogenous nucleation on aerosols in the upper troposphere. As noted by Fan et al. (2016), the dominant nucleation mechanism, or the competition between these two mechanisms, determines the aerosol impacts on cirrus clouds (Liu and
Penner 2005; Kärcher et al. 2006; Barahona and Nenes 2009; Gettelman et al. 2012). Higher INP concentrations can more efficiently compete for water vapor and inhibit homogeneous nucleation in regions where homogeneous nucleation dominates (e.g., Southern Hemisphere midlatitudes), thereby decreasing ice crystal number concentrations and increasing mean size. On the other hand, in regions already dominated by heterogeneous nucleation (e.g., polluted regions of the Northern Hemisphere), additional INP will lead to increased number concentrations and smaller ice crystal sizes.

Interestingly, Fan et al. (2016) suggest that the abundance of mineral dust—which is controlled by global weather patterns and may change significantly in future climates—can play a key role in modulating the competition between these two mechanisms. In addition to the roles played by atmospheric particles, cirrus formation and persistence depend on atmospheric dynamics at both small and large scales and on interactions with radiation.

Figure 11-30 demonstrates the expected aerosol impacts on in situ formed cirrus that prevail at higher altitudes than

![Schematic diagram of aerosol–orography–precipitation (AOP) interaction](image)

**FIG. 11-29.** Schematic diagram of aerosol–orography–precipitation (AOP) interaction for warm and cold phase orographic clouds with a (a) clean/pristine scenario, (b) polluted scenario, and (c) polluted scenario with high moisture content. For (b) and (c), + and − symbols are indicating an increase and decrease in precipitation over the windward or leeward side of the mountain. The +, −, or = symbols inside the circle indicate a change in net precipitation over the mountain. [Figure and caption from Choudhury et al. (2019). Reprinted with permission from Elsevier.]

**FIG. 11-30.** Total number concentration of ice crystals $n_i$ (y axis) formed in a cooling air parcel as a function of the vertical velocity $w$ (x axis) for an initial temperature of (left) 210 or (right) 230 K. The total number concentrations of heterogeneous ice nuclei freezing around a critical ice saturation ratio of 1.3 are indicated in the legend. The thick gray curves represent the pure homogeneous freezing cases. The curves are results of the parameterization; the symbols are taken from microphysical simulations performed with identical boundary conditions. [Figure and caption from Kärcher et al. (2006).]
liquid-origin cirrus (Wernli et al. 2016; Kramer et al. 2016), impacts that arise through particles acting as INPs as summarized in the review by Kärcher (2017). The calculations are parcel model simulations of the competition between homogeneous and heterogeneous nucleation described above, used to develop a parameterization of this process for global climate models. While an increase in CCN leads to an increase in the cloud droplet number concentration, the results in Fig. 11-30 show that an increase in INPs can increase or decrease ice crystal number concentrations compared with the homogeneous freezing limit (Kärcher et al. 2006), depending on the updraft velocity, temperature and the number concentration of INPs. The ice crystal number concentration approximately depends on the updraft velocity to the power of 3/2, which is much stronger than for cloud droplet number concentrations, where this dependence is only approximately 3/10 (Kärcher and Lohmann 2002). The importance of INPs for cirrus cloud formation needs to be unraveled from the impacts due to the small-scale variability in large-scale drivers, such as updraft velocities, that remains uncertain (Kärcher 2017). Since these can be competing effects, the impact of aerosols from volcanic eruptions on cirrus clouds remains inconclusive (Luo et al. 2002; Lohmann et al. 2003; Meyer et al. 2015). Also, in light of the geoengineering debate discussed below, it is important to know which fraction of the natural cirrus clouds forms due to homogeneous nucleation and thus could be altered by deliberate seeding with INPs (Mitchell and Finnegan 2009; Storelmo et al. 2013; Lohmann and Gasparini 2017).

j. Deep convective cloud systems

The impacts of aerosols on deep convective clouds and precipitation are perhaps the most complex of all, as they can potentially manifest through all of the microphysical processes described above. The Twomey effect suggests that higher concentrations of CCN lead to more numerous, smaller cloud droplets. The lifetime effect suggests that the development of precipitation is suppressed in the presence of increased aerosol concentrations, due to reduced cloud droplet sizes and collision coalescence efficiencies that suppress warm-rain processes (Squires and Twomey 1961; Warner and Twomey 1967; Warner 1968). This effect becomes increasingly complicated when it is extended to mixed-phase convective clouds, with widely varying results reported in the literature that examines the impacts of aerosols on convective storm systems.

The theoretical basis underpinning the impacts of aerosols on convective storm characteristics has rapidly evolved over the last decade, and not without controversy. The suppression of warm-rain formation in mixed-phase convective systems under enhanced aerosol concentrations allows additional cloud water to be lofted above the freezing level within these storm systems and hence made available to ice processes. Aerosol-induced changes to the cloud droplet size distributions of the lofted supercooled water would also impact ice processes. As such, it has been reported that the impacts of aerosol loading on the warm-rain process subsequently also impact mixed-phase processes including riming, melting, and cloud droplet freezing, all of which have implications for graupel and hail formation, as well as cloud morphology such as convective anvils. In addition to the numerous impacts of aerosols on mixed-phase microphysical processes, it has been hypothesized that variations in aerosol loading may impact the dynamical processes and features of convective storms. As the additional cloud water that is made available through the suppression of the warm-rain process is lofted above the freezing level and freezes, it releases latent heating. This additional heating enhances the buoyancy of the storm, and hence promotes stronger updrafts in more polluted conditions (Andreae et al. 2004; Khain et al. 2005; van den Heever et al. 2006), a process described as “convective invigoration” (Rosenfeld et al. 2008a). While this term is most often taken to mean stronger convective updrafts in more polluted conditions, convective invigoration has also been invoked to suggest higher cloud tops and enhanced precipitation rates (Fig. 11-31). More recently, “warm phase invigoration” has been suggested (Sheffield et al. 2015), in which enhanced bulk condensational heating within the warm phase of convective clouds enhances the updraft strengths in this region and may be more important than enhancements in the mixed-phase regions (Storer and van den Heever 2013; Li et al. 2013; Lebo 2014, 2018). This bulk condensational heating arises from the fact that for the same liquid water contents, a population composed of more numerous smaller droplets will have a larger total surface area on which to condense water and hence release latent heat.

The use of observational platforms to examine aerosol impacts on convective cloud characteristics spans more than 50 years. The impacts of aerosols on deep convective storms were initially examined through the use of in situ observational data and were rooted in part in the desire to understand the impacts of urban aerosols on downwind rainfall (Hobbs et al. 1970; Hindman et al. 1977a,b; Braham et al. 1981; Changnon 1981; Mather 1991; Borys et al. 1998; Ohashi and Kida 2002; Shepherd and Burian 2003; Jin et al. 2005; Shepherd 2005; van den Heever and Cotton 2007) and lightning (Orville et al. 2001; Williams et al. 2002; Steiger and Orville 2003). Since then, aerosol observations have been significantly advanced through the development of multisensor
(passive, active, visible, IR) airborne and space-based platforms including those on the A-Train (e.g., Kaufman and Nakajima 1993; Rosenfeld and Lensky 1998; Rosenfeld 1999, 2000; Rosenfeld and Woodley 2000; Andreae et al. 2004; Lin et al. 2006; Koren et al. 2005, 2010; Li et al. 2011; Heiblum et al. 2012; Storer et al. 2014). These multisensor studies have suggested that enhanced aerosol loading is associated with the presence of supercooled liquid water at greater altitudes, enhanced cold-rain and hail processes, higher cloud-top heights, and stronger updrafts.

The complexity of these systems, however, leads to the use of numerical models to develop understanding of aerosol-induced microphysical and dynamical processes, as well as providing a means to disentangle the role of environmental conditions and associated covariances. Rapid advancements in this arena are relatively recent, coinciding with the growth of computational capabilities. The impacts of aerosols on a wide range of different convective storm systems have been simulated, ranging from boundary layer cumulus and congestus (e.g., Li et al. 2013; Wall et al. 2014; Sheffield et al. 2015), through the more organized mesoscale systems of squall lines (e.g., Wang 2005; Tao et al. 2007, 2012; Fan et al. 2007, 2017; Li et al. 2009; Seigel et al. 2013; Lebo and Morrison 2014; Marinescu et al. 2017) and, more recently, mesoscale convective complexes (Clavner et al. 2018a,b) and severe convective storms including supercells and tornados (e.g., Lerach et al. 2008; Snook and Xue 2008; Seifert and Beheng 2006; Khain and Lynn 2009; Storer et al. 2010; Lebo and Seinfeld 2011; Morrison 2012; Lerach and Cotton 2012; Lebo et al.; Kalina et al. 2014; Lerach and Cotton 2018) and tropical cyclones (e.g., Dunion and Velden 2004; Rosenfeld et al. 2007; H. Zhang et al. 2007, 2009; Jenkins and Pratt 2008; Cotton et al. 2007; Khain et al. 2010; Carrió and Cotton 2011; Rosenfeld et al. 2011; Cotton et al. 2012; Rosenfeld et al. 2012; Herbener et al. 2014, 2016). While a few studies have examined the impacts of aerosols on the overall response of a

![FIG. 11-31. Schematic illustration of the differences in cloud-top height, cloud fractions, and cloud thickness for the storms in clean and polluted environments. Red dots denote cloud droplets, light blue dots represent raindrops, and blue shapes are ice particles. In the polluted environment, convective cores detrain larger amounts of cloud hydrometeors of much smaller size, leading to larger expansion and much slower dissipation of stratiform/anvil clouds resulting from smaller fall velocities of ice particles because of much reduced sizes. Therefore, the larger cloud cover, higher cloud-top heights, and thicker clouds are seen in the polluted storm after the mature stage. [Figure and caption from Fan et al. (2013).]
collection of different storm types (Seifert and Beheng 2006; van den Heever et al. 2011), most of the aerosol–cloud deep convection studies in the literature have examined individual storm systems, often focusing on specific morphological characteristics, including surface precipitation, updraft cores, anvil properties, cold pools, and/or lightning. We present a few illustrative results for each of these characteristics.

The impact of aerosols on convective precipitation is still not well understood, and increases in aerosol number concentrations have been found to produce a range of responses in surface rainfall: systematic increases (e.g., Phillips et al. 2002; Khain and Pokrovsky 2004, 2005; Teller and Levin 2006), systematic decreases (Wang 2005; van den Heever et al. 2006; Fan et al. 2007; Lee et al. 2009), or little change (see Table 3 of Tao et al. 2005; Teller and Levin 2006), systematic decreases (Khain et al. 2005; van den Heever et al. 2006; Fan et al. 2007; Lee et al. 2009), or little change (see Table 3 of Tao et al. 2012). The differences in simulated surface precipitation trends with increasing aerosol concentrations have been attributed to modulation by various environmental factors.

Aerosol-induced enhanced updraft strengths within the mixed-phase regions of deep convection were initially identified both in observational studies (Rosenfeld and Woodley 2000; Andreae et al. 2004; Koren et al. 2005; Lin et al. 2006) and in cloud-resolving simulations (Khain et al. 2005; van den Heever et al. 2006) and were attributed to convective invigoration from mixed-phase processes. More recently, however, warm-phase invigoration was suggested to be of equivalent or greater importance (Storer and van den Heever 2013; Li et al. 2013; Lebo 2014, 2018). Downdrafts also appear to be impacted by the presence of aerosols (Khain et al. 2005), with impacts for convective cold pools.

While more numerous, smaller cloud droplets are formed under enhanced aerosol loading, it has been found in both modeling (Altaratz et al. 2007; Storer et al. 2010; Storer and van den Heever 2013) and observational studies (Berg et al. 2008; May et al. 2011) that the associated raindrop population is composed of less numerous drops that are larger in size. This outcome arises from enhanced collision–coalescence efficiencies in the presence of the additional cloud water available in more polluted conditions. For precipitation comprising fewer raindrops that are larger in size, bulk evaporation rates will be reduced, subsequently resulting in warmer downdrafts, and hence warmer and weaker cold pools (van den Heever and Cotton 2007; Storer et al. 2010). This feedback has implications for storm strength and maintenance, as well as for new storm initiation. However, other studies have suggested that raindrop populations may instead comprise smaller droplets and hence that cold pools may be stronger in polluted environments (Tao et al. 2007; Lee et al. 2009). It has been speculated that these differences may be attributed to environmental modulation of the storm life cycle.

The generation of additional cloud water with a shift to smaller cloud droplet sizes, and lifting of this liquid water to the mixed and ice phases of convective systems (e.g., Heikenfeld et al. 2019), has the potential to significantly impact the ice processes within the anvil, including those processes impacting snow, graupel, and hail formation (e.g., aggregation, riming, melting, and homogeneous freezing). A number of modeling studies have focused on aerosol indirect effects on the upper-level convective anvils, of particular interest given the role of these high clouds in radiation budgets, as mentioned above. Some of these studies have found a monotonic increase in anvil ice water content with increasing aerosol number concentrations (van den Heever et al. 2006; Carrió et al. 2007) while others demonstrate nonmonotonic responses (Fan et al. 2007; Li et al. 2008; Carrió et al. 2010; Loftus and Cotton 2014). An increase in anvil lifetimes has been observed in more polluted conditions through the generation of persistent small ice particles with smaller fall speeds (Khain et al. 2005; Carrió et al. 2007). Anvil size has, however, been found to either increase (Khain et al. 2005; Fan et al. 2010a,b) or decrease (van den Heever et al. 2006) in the presence of increased aerosol loading. Furthermore, aerosol-induced changes to ice crystal size and number concentrations have also been observed to impact anvil cloud radiative properties and associated radiative forcing (Carrió et al. 2007; Lee et al. 2009). The impact of aerosols on the anvil properties appears to be related to the effect that aerosols have on the riming of cloud water, which has been found to increase in some cases and decrease in others. In recent simulations of a mesoscale convective system, Saleeby et al. (2016) demonstrated an increase in the rime collection rates of cloud water under increased aerosol loadings, and hence in that study, less cloud water mass reached the anvil. However, the ice mass was composed of more numerous smaller ice crystals, resulting in larger anvils, enhanced albedo, decreased cloud-top cooling, and a reduced net radiative flux, thereby demonstrating a potential aerosol-induced warming effect of squall lines forming under polluted conditions.

Given the apparent impact of aerosols on vertical updrafts, supercooled water, riming, and the depth of the mixed-phase cloud, it has been suggested that aerosols could play a significant role in modifying lightning production and frequency (Rosenfeld and Woodley 2000; Williams et al. 2002, 2005). Enhanced lightning flash densities have been observed to occur in association with enhanced urban, volcanic, and other aerosols (e.g., Orville et al. 2001; Steiger and Orville...
Variations in environmental conditions have been found to significantly impact the trends in aerosol indirect effects on convective clouds. Tao et al. (2007) demonstrated varying responses to aerosol loading in a moist maritime environment as compared with a drier midlatitude environment. Khain et al. (2008) and Khain (2009) concluded from a number of numerical experiments that high environmental relative humidity leads to an increase in precipitation. Vertical wind shear also appears to play a role in modulating the impacts of increased aerosol concentrations on precipitation, with convection suppression (enhancement) under strong (weak) wind shear conditions (Fan et al. 2009), and CAPE is also a predominant factor in that aerosol impacts are less in conditions of high environmental CAPE compared with lower CAPE (Storer et al. 2010). Aerosol impacts on convective storms can also be affected by aerosol type. Both observational (Rosenfeld and Nirel 1996) and modeling studies (van den Heever et al. 2006; Ekman et al. 2007) have examined the impacts of dust serving as CCN, GCCN, and/or INP on deep convection and have argued that the enhancement in precipitation from dust serving as GCCN and INP can offset and/or modulate the suppression of precipitation by dust serving as CCN. The vertical profile of aerosols has been found to significantly impact the location of droplet nucleation and the initiation of subsequent microphysical processes (Fridlind et al. 2004; Marinescu et al. 2017). Finally, aerosol indirect effects also vary as a function of storm type (Seifert and Beheng 2006; Khain et al. 2008; van den Heever et al. 2011; Grant and van den Heever 2015).

6. Climate impacts of aerosols

In the discussions above we have noted that interest in both anthropogenic aerosol impacts on weather and biogeochemical cycles, and the role of aerosols in atmospheric composition and processes, existed before the advent of the twentieth century, but scientific studies were accelerated in the past half-century due to rapid advancements in the understanding of aerosol–cloud–climate interactions and in a sense of urgency regarding the potential for significant and long-term effects on Earth’s climate from greenhouse gas emissions, as moderated by aerosols. In response to international interest, in 1988 the World Meteorological Organization (WMO) and the United Nations Environment Program (UNEP) created the Intergovernmental Panel on Climate Change (IPCC) “to provide policymakers with regular assessments of the scientific basis of climate change, its impacts and future risks, and options for adaptation and mitigation” (http://www.ipcc.ch). The First Assessment Report was delivered in 1990, and the Sixth Assessment Report is due to be finalized in 2022. The IPCC has served to not only gather and review available studies, but also to help spur model development (Randall et al. 2019). The Assessment Reports document the evolution of scientific understanding as models have become more powerful and have included more detailed representations of key atmospheric processes, and as many more data from observing systems have become available to constrain modeled relationships and outcomes.

The summary table of radiative forcing estimates between preindustrial times representing 1750 and the present-day climate in 2011 included in the Fifth Assessment Report is shown in Fig. 11-32 (Stocker et al. 2013). In this figure, aerosol forcing (the difference in radiative flux between present day and preindustrial conditions) has been attributed to specific chemical components, with black carbon estimated to contribute a net warming in contrast to the cooling associated with anthropogenic sulfate; the uncertainties on the sign and magnitude of the overall effect are large. Similarly, significant uncertainty remains in the global net forcing due to aerosol impacts on clouds, as discussed further below.

In the following, we review recent progress in the model treatments and estimates of aerosol direct and indirect effects on climate.

a. Direct effects

Aerosol particles influence climate by scattering and absorption of radiation. The degree to which aerosol particles affect the radiative budget is proportional to their amount in the atmosphere and to their ability to extinguish radiation. The extinction of radiation is the sum of the scattering and absorption of radiation. It is proportional to the aerosol mass mixing ratio and is a function of particle size and wavelength. Extinction is usually defined in terms of an extinction coefficient specific to each aerosol type. Aerosol radiative properties can vary greatly, depending on their scattering and absorbing properties. These depend on their shapes, sizes (larger particles absorb more and scatter less), chemical compositions and mixing states. In atmospheric science it has been proven useful to divide the extinction coefficient by the aerosol density to obtain the mass extinction efficiency (MEE; m² kg⁻¹).

The scattering and absorption of solar radiation by aerosol particles reduces the amount of solar radiation
absorbed at Earth’s surface, especially over dark regions such as oceans. The net effect of all aerosol particles (from natural and anthropogenic sources) on the shortwave radiation at the top-of-the-atmosphere is also negative since aerosol scattering prevails over absorption for most aerosol types. The effect of aerosol particles on the net radiation at the top-of-the-atmosphere is commonly referred to as radiative forcing (RF).

Larger particles (in the micrometer size range, e.g., mineral dust) have a significant positive RF in the longwave mainly due to their absorptive properties, while their longwave scattering is typically of minor importance. For instance, stratospheric sulfate aerosol particles exhibit a net negative RF as long as they are smaller than 2 μm, but larger particles have a net positive RF at longer wavelengths (Lacis et al. 1992). In the global annual mean the positive longwave RF of aerosol particles is smaller than the negative shortwave RF, so that the aerosol RF is negative (see Fig. 11-32).

The scattering of solar radiation by aerosol particles is dominated by Mie scattering and is largest for aerosol particles that have diameters comparable with the wavelengths of the electromagnetic radiation with which they interact. For the scattering of visible light, this corresponds to accumulation mode aerosol particles. For some aerosol particles, such as black carbon, the absorption of solar radiation dominates over the scattering of it.

Mass extinction values are not well known and different climate models use different values. The median values from 20 global climate models are 8.9 m² g⁻¹ for black carbon (range between 5.3 and 18.9 m² g⁻¹), 8.5 m² g⁻¹ for sulfate (range between 4.2 and 18 m² g⁻¹), 5.7 m² g⁻¹ for particulate organic matter (range between

<table>
<thead>
<tr>
<th>Emitted compound</th>
<th>Resulting atmospheric drivers</th>
<th>Radiative forcing by emissions and drivers</th>
<th>Level of confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO₂</td>
<td>CO₂</td>
<td>1.66 [1.33 to 2.03]</td>
<td>VH</td>
</tr>
<tr>
<td>CH₄</td>
<td>CO₂ H₂O₉ O₃ CH₄</td>
<td>0.97 [0.74 to 1.20]</td>
<td>H</td>
</tr>
<tr>
<td>Halo-carbons</td>
<td>O₃ CFCs HCFCs</td>
<td>0.16 [0.01 to 0.35]</td>
<td>H</td>
</tr>
<tr>
<td>N₂O</td>
<td>N₂O</td>
<td>0.17 [0.13 to 0.21]</td>
<td>VH</td>
</tr>
<tr>
<td>Anthropogenic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>CO₂ CH₄ O₃</td>
<td>0.23 [0.16 to 0.30]</td>
<td>M</td>
</tr>
<tr>
<td>NMVOC</td>
<td>CO₂ CH₄ O₃</td>
<td>0.10 [0.05 to 0.15]</td>
<td>M</td>
</tr>
<tr>
<td>NO₂</td>
<td>Nitrate CH₄ O₃</td>
<td>-0.15 [-0.34 to 0.03]</td>
<td>M</td>
</tr>
<tr>
<td>Aerosols and precursors</td>
<td>Mineral dust Sulphate Nitrate Organic carbon Black carbon</td>
<td>-0.27 [-0.77 to 0.23]</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>Cloud adjustments due to aerosols</td>
<td>-0.55 [-1.33 to -0.06]</td>
<td>L</td>
</tr>
<tr>
<td>Natural</td>
<td>Changes in solar irradiance</td>
<td>0.05 [0.00 to 0.10]</td>
<td>M</td>
</tr>
</tbody>
</table>

**Total anthropogenic RF relative to 1750**

<table>
<thead>
<tr>
<th>Year</th>
<th>Radiative forcing relative to 1750 (W m⁻²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>2.29 [1.13 to 3.33]</td>
</tr>
<tr>
<td>1980</td>
<td>1.26 [0.64 to 1.86]</td>
</tr>
<tr>
<td>1950</td>
<td>0.57 [0.29 to 0.85]</td>
</tr>
</tbody>
</table>
3.2 and 11.4 m² g⁻¹, 3 m² g⁻¹ for sea salt aerosols (range between 0.9 and 7.5 m² g⁻¹), and 0.95 m² g⁻¹ for mineral dust (range between 0.5 and 2.1 m² g⁻¹) (Kinne et al. 2006). The MEEs for sulfate and carbonaceous aerosols are typically larger than those for sea salt and mineral dust because of their generally smaller sizes, with a larger fraction of accumulation mode particles than for sea salt and mineral dust.

Since preindustrial times, anthropogenic aerosol emissions have caused a considerable perturbation to the radiation balance. Between 1750 and 2014, SO₂ emissions have increased from 0.5 to 111.7 Tg yr⁻¹, those of particulate organic matter from 3.1 to 27.6 Tg yr⁻¹ and those of black carbon from 0.5 to 8 Tg yr⁻¹ (Hoesly et al. 2018).

The enhanced scattering and absorption of radiation by anthropogenic aerosols is known as the direct aerosol effect or, more recently, as the radiative forcing due to aerosol–radiation interactions (RFₐₐᵢ; Fig. 11-33). It is estimated in terms of a global annual mean change since preindustrial times. Our knowledge about the preindustrial atmosphere is rather limited; therefore, 1850 is frequently assumed as the preindustrial comparative year, due to more availability of data, although some anthropogenic activity had already influenced the atmosphere by that time. Therefore, 1750 is the more appropriate reference year (Myhre et al. 2013).

The first attempt to estimate RFₐₐᵢ dates back to Charlson et al. (1990), who focused on sulfate aerosols as the best studied aerosol species at that time because of international efforts to understand and mitigate acid rain. They used a simple box model calculation to estimate that RFₐₐᵢ is comparable to the forcing of anthropogenic carbon dioxide (at the time of their study) and amounts to approximately −1.6 W m⁻². Since then, many estimates of RFₐₐᵢ were conducted and not only sulfate aerosols, but also black carbon, particulate

---

**FIG. 11-33.** Schematic of RF and ERF due to aerosol–radiation interactions (ari) and to aerosol–cloud interactions (aci). The terminology used here follows that used in Boucher et al. (2013). The thick arrows depict the incoming solar radiation, and the thin arrows the reflected and transmitted solar radiation. The gray and black squares in the upper figure denote scattering and absorption by aerosol particles, and the gray disks in the upper and lower figures denote larger and smaller cloud droplets, whereby smaller gray disks refer to smaller cloud droplets. [Figure and caption from Lohmann et al. (2016a). © Ulrike Lohmann, Felix Lüööd, and Fabian Mahrt 2016. Reproduced with permission of The Licensor through PLSclear.]
organic matter, sea salt, and mineral dust are considered in most current global aerosol–climate models. Nitrate aerosols are considered only in a minority of global aerosol–climate models because treating them requires including gas-phase chemistry that is lacking in most climate models. The most recent summary in the Fifth Assessment Report of the IPCC (AR5) is based on a combination of global aerosol–climate models and observation-based methods. It yields an estimate of \( \text{RF}_{\text{ari}} \) of \(-0.35 \pm 0.03 \) W m\(^{-2}\) (Myhre et al. 2013), which is less than 25% of the first estimate by Charlson et al. (1990). The AR5 value of \( \text{RF}_{\text{ari}} \) is smaller than its assessment in AR4 of \(-0.5 \) W m\(^{-2}\) (Forster et al. 2007) because of a reevaluation of aerosol absorption.

\( RF_{\text{ari}} \) can be broken down into the modeled contributions from the individual aerosol species, as shown in Fig. 11-32, but this is more uncertain as can be appreciated from the differences in MEE and aerosol burdens (Kinne et al. 2006) between different climate models. For the time period spanning 1750 to 2011, \( RF_{\text{ari}} \) has been estimated to amount to \(-0.4 \pm 0.2 \) W m\(^{-2}\) for sulfate aerosols, \(0.4 \pm 0.2 \) W m\(^{-2}\) for black carbon aerosols from fossil fuel and biofuel, \(-0.09 \pm 0.03 \) W m\(^{-2}\) for primary organic aerosols from fossil fuel and biofuel, \(-0.09 \pm 0.03 \) W m\(^{-2}\) for aerosols associated with biomass burning, \(-0.03 \pm 0.02 \) W m\(^{-2}\) for secondary organic aerosols, and \(-0.11 \pm 0.03 \) W m\(^{-2}\) for nitrate aerosols, and \(-0.1 \pm 0.1 \) W m\(^{-2}\) for dust aerosols (Myhre et al. 2013). The large uncertainty for black carbon arises from insufficient information about its emissions, its vertical distribution and its removal rate, which in turn strongly rely on its poorly known interaction with clouds (Koch et al. 2011; Bond et al. 2013). Black carbon is the only aerosol species that causes a positive radiative forcing due to its high absorption cross section. However, black carbon rarely occurs in isolation from other emitted species and little is known as to how its mixing with other substances influences \( \text{RF}_{\text{ari}} \) (Bond et al. 2013).

In addition to the radiative forcing, fast adjustments occur on time scales that are much shorter than the time scales associated with those of the long-lived greenhouse gases. Therefore, AR5 adopted the notion of an effective radiative forcing (ERF) that considers these fast adjustments (Fig. 11-33). In terms of aerosol–radiation interactions, these include for instance a change in static stability due to the absorption of solar radiation by black carbon and a possible associated change in cloud cover. A calculation of \( \text{RF}_{\text{ari}} \) is done by performing two radiative transfer calculations in a GCM simulation for present-day conditions. The first radiative transfer calculation includes all aerosol species and everything else, while for the second one everything else remains at present-day conditions, except for the aerosol specie(s) in question that is set to its preindustrial value. A calculation of \( \text{ERF}_{\text{ari}} \) on the other hand, requires two separate climate model simulations, one with preindustrial aerosol emissions and one with present-day aerosol emissions. In the present-day simulation, the meteorology will be different from the one at preindustrial times due to the interactions of anthropogenic aerosols with the climate system giving rise to various cloud adjustments.

A summary of the most recent estimates of \( \text{ERF}_{\text{aci}} \) from the models that participated in the phase 5 of the Coupled Model Intercomparison Project (CMIP5)/Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP) can be found in IPCC AR5. \( \text{ERF}_{\text{ari}} \) from these models amounts to \(-0.45 \) W m\(^{-2}\) (Boucher et al. 2013) and is slightly more negative than \( \text{RF}_{\text{aci}} \) of \(-0.35 \) W m\(^{-2}\) (Stocker et al. 2013) because of fast adjustments in clouds in response to aerosol absorption. These adjustments are mainly caused by the absorption of solar radiation by black carbon and are also referred to as the semidirect effect (Fig. 11-33: Hansen et al. 1997; Ackerman et al. 2000; Lohmann and Feichter 2001). Whereas earlier studies concluded that the semidirect effect is positive because of the reduction in cloudiness due to the warmer temperatures, newer studies found this effect to be negative and to partly offset \( \text{RF}_{\text{ari}} \) from black carbon (Stjern et al. 2017). The negative semidirect effect is caused by strong temperature adjustments at altitudes above 400 hPa where increased stabilization and reduced cloud cover cause larger emissions of longwave radiation to space. However, the vertical distribution of aerosol absorption is not well known, leading to uncertainties in these adjustments.

b. Indirect effects

In addition to \( \text{ERF}_{\text{aci}} \), climate models have attempted to include effects of aerosol particles on the microphysical structure of clouds through their activities as CCN and INPs. To recap the impacts on cloud systems discussed in prior sections, and to demonstrate how these interactions are conceptualized for parameterization on large-scale models, we note that an increase in anthropogenic aerosol particles leads to an increase in CCN, which in turn increases the (initial) cloud droplet number concentration. Assuming that the cloud liquid water content remains the same, this increase in the cloud droplet number concentration is associated with a decrease in the cloud droplet size (as seen in the ship track in Fig. 11-27). A cloud that consists of more but smaller cloud droplets has a higher overall surface area and hence scatters more solar radiation back to space—that is, the Twomey effect. The first climate model studies estimating the global impact of the Twomey
effect, expressed as RF due to aerosol–cloud interactions ($\text{RF}_{\text{aci}}$) in warm clouds resulting from an increase in CCN, used sulfate aerosols as a surrogate for all anthropogenic aerosols. These studies were published 20 years after Twomey’s seminal 1974 work (Boucher and Lohmann 1995; Jones et al. 1994) and obtained $\text{RF}_{\text{aci}}$ ranging from $-1$ to $-1.3 \text{ W m}^{-2}$.

Similar to $\text{RF}_{\text{ari}}$, $\text{RF}_{\text{aci}}$ is a hypothetical construct that is obtained by calling the radiative transfer algorithm twice, keeping everything but the change in cloud droplet number concentration and size the same at present-day conditions. In reality such an effect cannot be observed because a change in the cloud droplet number concentration and size affects the rate of precipitation formation, evaporation of cloud droplets, and cloud-top cooling. Therefore, as with the evolution in thinking around $\text{RF}_{\text{ari}}$, the modeling community has moved away from $\text{RF}_{\text{aci}}$ and modeling studies now include fast adjustments. In the case of aerosol–cloud interactions, these fast adjustments are associated with changes in the lifetime of the cloud, its phase, and its vertical extent (Fig. 11-33 and Boucher et al. 2013).

The importance of the cloud lifetime effect on a global scale remains controversial. Most GCMs include a cloud lifetime effect because their autoconversion rates of cloud droplets to form raindrops depend inversely on the cloud droplet number concentrations, that is, an increase in the cloud droplet number concentration slows down drizzle formation and leads to an increase in liquid water path. While such an increase in liquid water path is found in ship tracks crossing open-cell regimes, in closed-cell clouds, little change in cloud properties was observed (Christensen and Stephens 2011; Chen et al. 2015). Recently, analyses from clouds downwind of the 2014–15 Holohraun eruption (Malavelle et al. 2017), which emitted SO$_2$ amounts corresponding to almost twice the annual emissions of Europe, showed an area-mean decrease of about 1 $\mu$m in the effective radius in October 2014 with little increase in area-mean liquid water path. To put this number into perspective, a global mean decrease in effective radius of about 1 $\mu$m corresponded to a global mean $\text{RF}_{\text{aci}}$ of $-1 \text{ W m}^{-2}$ in the study by Boucher and Lohmann (1995). The Holohraun study is a convincing example of the importance of buffering effects in the climate system (Stevens and Feingold 2009). However, while this study demonstrates that compensating effects can be observed on regional to global scales, we note that local changes in liquid water paths may yet be large and as such, may have significant impacts on precipitation and other societally relevant phenomena at those scales.

In addition to adjustments in the warm phase, the cloud phase, cloud height, and cloud cover can also change. A change in the cloud phase can be initiated if anthropogenic activity changes the number concentration and/or the properties of the INPs. If anthropogenic activity leads to more INPs in the mixed-phase temperature range between 0°C and $-38^\circ$C, then more ice crystals form, which speeds up the WBF process (see section 5) and causes the mixed-phase cloud to glaciate. Because stratiform clouds containing ice are shorter-lived than stratiform liquid clouds, this so-called glaciation indirect aerosol effect (Lohmann 2002) partly offsets $\text{ERF}_{\text{aci}}$ in the warm phase. However, anthropogenic activity primarily leads to an increase in soluble aerosols, such as sulfates, and not necessarily to an increase in INP. Further, if sulfuric acid condenses on potential INPs such as mineral dust or soot particles, it may cover their active sites, resulting on the contrary in a deactivation indirect aerosol effect (Hoose et al. 2008; Storelvmo et al. 2008). Which effect prevails is not yet clear.

On the scale of ship tracks, satellite studies suggest that the glaciation indirect effect prevails (Christensen et al. 2014). Even in the absence of anthropogenic INP, an increase in CCN, for instance due to ship exhaust, can lead to more cloud-top cooling and hence freezing in Arctic mixed-phase clouds (Possner et al. 2017). The glaciation indirect effect can also be initiated by anthropogenic dust emissions or to a weaker extent also by smoke aerosols (Tan et al. 2014).

Anthropogenic aerosols could change cloud heights via convective invigoration (Rosenfeld et al. 2008a, 2014). To recap salient points from the discussion in the preceding section, in pristine deep convective clouds, cloud droplets grow by collision–coalescence to raindrop size and leave the cloud. In polluted clouds, where growth by collision–coalescence is suppressed, cloud droplets are transported to higher levels with colder temperatures, where they can freeze. The freezing temperature depends on the size of the cloud droplets, such that smaller droplets freeze at colder temperatures (both homogeneously and heterogeneously) than larger drops because the chance of forming a critical ice embryo or of a drop containing an INP increases with increasing droplet size. However, in convective clouds, two effects act in opposition. More condensate in the polluted cloud reduces the buoyancy due to increased water loading. On the other hand, freezing releases latent heat aloft and invigorates these cloud systems (Koren et al. 2005; Rosenfeld et al. 2008a), especially if hydrometeors leave the cloud. Also, invigoration of warm convective clouds may occur through faster condensation (and thus faster rates of latent heating) in polluted clouds (Koren et al. 2014). If pollution also increases the ice crystal number concentration in
polluted clouds, this would increase the lifetime and coverage of the stratiform anvil in the mature and dissipating stages of deep convective clouds due to the slower sedimentation of the smaller ice crystals (Fan et al. 2013). However, separating aerosol effects from dynamical effects on deep convective clouds, and representing these effects in global-scale models, remains challenging (Grabowski 2018).

\( \text{ERF}_{\text{aci}} \) from all bottom-up climate model estimates (imposing no constraints from the observed temperature record) is \(-1.5 \text{ W m}^{-2}\) (Boucher et al. 2013). This includes climate models that treat aerosol effects on liquid water and mixed-phase clouds, but disregards estimates of aerosol effects on cirrus clouds because not enough were published at the time of the IPCC AR5 report for a thorough assessment. As discussed in the prior section, aerosols affect cirrus clouds through heterogeneous freezing on INP and homogeneous freezing of solution droplets. While the homogeneous nucleation of cirrus clouds is only rarely limited by the availability of solution droplets, heterogeneous nucleation is often limited by the scarcity of INP (Kärcher 2017). However, our understanding of the dominant cirrus formation mechanism in the present atmosphere is still insufficient (Cziczo et al. 2013; Gasparini et al. 2018), which poses uncertainties for simulating the anthropogenic effect on cirrus clouds in climate models. If soot is assumed to be an efficient INP and the pristine cirrus clouds in preindustrial times are assumed to form by homogeneous nucleation, then the polluted cloud would nucleate fewer ice crystals at warmer temperatures and lower supersaturations. These fewer ice crystals will grow to larger sizes and leave the cloud, further reducing its optical depth. While such a cloud reflects less solar radiation, the dominant effect is the higher emission of longwave emissions to space, so that \( \text{ERF}_{\text{aci}} \) is negative from \(-0.3\) to \(-0.4 \text{ W m}^{-2}\) (Penner et al. 2009). If, on the contrary, the polluted cirrus clouds consist of more ice crystals, \( \text{ERF}_{\text{aci}} \) will have the opposite sign. In the study by Gettelman et al. (2012), who investigated the impact of all anthropogenic aerosols on cirrus clouds, they found the latter effect to dominate, and assessed \( \text{ERF}_{\text{aci}} \) of cirrus clouds to be \(0.3 \pm 0.1 \text{ W m}^{-2}\) using different methods and models. In their study, this effect represented a 20% offset of the simulated shortwave \( \text{ERF}_{\text{aci}} \) of ice and liquid clouds of \(-1.6 \text{ W m}^{-2}\).

\( \text{c. Combined effects} \)

The distinction between \( \text{ERF}_{\text{aci}} \) and \( \text{ERF}_{\text{ari}} \) is not straightforward, because changes in the clear sky can also affect cloud formation and changes in cloud micro and macrophysics can also affect aerosol concentrations in the clear sky through washout by precipitation. Therefore, in IPCC AR5 the scientific community focused on the assessment of the overall \( \text{ERF}_{\text{ari+aci}} \) (Boucher et al. 2013), which was estimated as \(-0.9 \text{ W m}^{-2}\) with a range between \(-1.9\) and \(-0.1 \text{ W m}^{-2}\). This expert judgement was obtained by considering only those models that include a more complete or consistent treatment of aerosol–cloud interactions by including effects on mixed-phase or convective clouds, only taking the latest estimate from a given climate model, and equally weighting estimates of \( \text{ERF}_{\text{ari+aci}} \) that considered satellite estimates and those that were derived from bottom-up climate model studies. The AR5 range of \( \text{ERF}_{\text{ari+aci}} \) is slightly less negative than the previous expert judgement solicitation by Morgan et al. (2006), which ranged between \(-2.1\) and \(-0.25 \text{ W m}^{-2}\).

Estimates of \( \text{ERF}_{\text{ari+aci}} \) that include satellite data yield an average value of \(-0.8 \text{ W m}^{-2}\) (Boucher et al. 2013). Satellite data have the problem that aerosols and clouds cannot be detected concurrently, which introduces some uncertainty. On the other hand, GCMs that estimate \( \text{ERF}_{\text{ari+aci}} \) arrive at an average of \(-1.3 \text{ W m}^{-2}\) (Boucher et al. 2013). Even though IPCC AR5 only considered those studies that include a more complete or consistent treatment of aerosol–cloud interactions by including effects on mixed-phase or convective clouds, there is still reason to believe that GCMs probably do not account for buffering effects (Stevens and Feingold 2009) and that their coarse resolution overestimates the radiative forcing (Possner et al. 2016). Therefore, the expert judgement in IPCC AR5 was to downplay the importance of the GCM estimates and to endorse a central value of \(-0.9 \text{ W m}^{-2}\) for \( \text{ERF}_{\text{ari+aci}} \).

Some problems, however, remain: If we assume that \( \text{ERF}_{\text{ari}} \) and \( \text{ERF}_{\text{aci}} \) should be additive, then \( \text{ERF}_{\text{ari+aci}} \) should be \(-2 \text{ W m}^{-2}\) (\( \text{ERF}_{\text{ari}} = -0.45 \text{ W m}^{-2} + \text{ERF}_{\text{aci}} = -1.5 \text{ W m}^{-2}\)). This is not the case because as stated above \( \text{ERF}_{\text{ari+aci}} \) amounts to only \(-1.3 \text{ W m}^{-2}\). This discrepancy partly arises because these numbers are averages over all published best estimates, but normally models report either \( \text{ERF}_{\text{aci}} \) and \( \text{ERF}_{\text{ari}} \) or \( \text{ERF}_{\text{ari+aci}} \). In other words, the estimates of \( \text{ERF}_{\text{aci}} \) and \( \text{ERF}_{\text{ari}} \) or \( \text{ERF}_{\text{ari+aci}} \) originate from different models. This comparison also demonstrates that \( \text{ERF}_{\text{aci}} \) estimates alone are less meaningful.

Another approach to obtain \( \text{ERF}_{\text{ari+aci}} \) is to deduce it from those coupled atmosphere–ocean GCMs that both participated in CMIP5 and that matched the observed warming. This yields an estimate of \( \text{ERF}_{\text{ari+aci}} = -0.8 \pm 0.9 \text{ W m}^{-2} \) (Forster et al. 2013), which corresponds well to the IPCC AR5 estimate of \(-0.9 \text{ W m}^{-2}\). Likewise, inverse methods, which start from the observed temperature record and the known increase in greenhouse gases, exist but rely on assumptions of the ocean heat uptake.
and of the transient climate sensitivity, that is, how much of the warming is already realized. An early study of such an inverse estimate of ERFaci by Knutti et al. (2002) that is compatible with the observed temperature record brackets ERFaci between 0 and $2.17 \text{ W m}^{-2}$ (Knutti et al. 2002). IPCC AR5 arrived at an expert judgement of a range for ERFari from inverse estimates between 0 and $2.15 \text{ W m}^{-2}$ (Bindoff et al. 2013).

What is new since AR5? The average of all published ERFari estimates as of 2017 from bottom-up model studies and those involving satellite data is $2.13 \text{ W m}^{-2}$ (Lohmann 2017b). This average does include our process understanding but also includes older and sometimes outdated estimates. A recent study by Nazarenko et al. (2017) investigates the impact of climate feedbacks on ERFari. The increase in cirrus clouds in the warmer climate tends to reduce ERFari down to $0.3$ and $0.74 \text{ W m}^{-2}$, respectively, depending on the aerosol scheme employed.

A summary of RFari, ERFari, RFaci, ERFaci, RFari+aci, and ERFari+aci estimates from the earliest estimates to the newest ones is shown in Fig. 11-34. From this representation, we note that the early estimates of RFari and ERFari in which sulfate was taken as a surrogate for all anthropogenic aerosols are higher than subsequent estimates, where multiple anthropogenic species were taken into account. Estimates of RFari from the CMIP5 models that are designed to match the observed
temperature record are generally smaller because a large aerosol forcing seems inconsistent with the observed warming (e.g., Stevens 2015). On the contrary, observational-based estimates of RF_{ari} are larger. Understanding this difference is ongoing work.

Again referring to Fig. 11-34, RF_{aci} is more negative than RF_{ari} and the estimates are more scattered because they also include uncertainties associated with the simulation of clouds and with parameterizations of aerosol–cloud interactions. Here the observational-based estimates are of smaller magnitude largely because satellite data have the problem of not being able to simultaneously retrieve clouds and aerosols.

The ordering of data within each category in Fig. 11-34 reflects the publication dates of the various studies. We note that there is no tendency for ERF_{ari+aci} estimates to become larger or smaller over time, except for those estimates that involve convective clouds. Thus, the most recent estimates of ERF_{ari+aci} in each category are from around $-1$ to $-1.5 \, \text{W m}^{-2}$, not so different from the earliest estimates of the sum of RF_{ari} and RF_{aci} using sulfate as a surrogate for all anthropogenic aerosols.

Mülenstädt and Feingold (2018) review recent progress in establishing ERF_{aci} for warm clouds, and suggest new approaches to accelerate progress in this key area, including studying the covariability of aerosol and meteorological drivers in order to narrow the relevant parameter space. Recent work on aerosol effects on mixed-phase clouds are reviewed by Lohmann (2017b); these studies and those focusing on aerosol effects on convective and cirrus clouds are less numerous than those for warm clouds. An update of the “highlighted” GCMs in IPCC AR5, that is, those that include a more complete or consistent treatment of aerosol–cloud interactions by including effects on mixed-phase or convective clouds, remains at ERF_{ari+aci} of $-1.3 \, \text{W m}^{-2}$ (Lohmann 2017a). As pointed out by Heyn et al. (2017), a more negative ERF_{ari+aci} is conceivable because compensating effects on the longwave radiation are probably smaller than assumed in IPCC AR5. However, for the time being, the effect of aerosols on mixed-phase and cirrus clouds remains associated with large uncertainty and is an area in which future work is needed. Likewise, the impact of horizontal resolution on the parameterizations of aerosol–cloud interactions and ERF_{ari+aci} (e.g., Possner et al. 2016) needs to be better understood and accounted for.

### 7. Applied microphysics

As the connections between cloud microphysics, precipitation formation, and the atmospheric aerosol were uncovered, attention turned to the possibility that those pathways could be manipulated for desired outcomes—for example, avoiding hail formation or ending a drought, bringing credible science behind humanity’s long-held desire to control the weather (Fleming 2010). Given the strong connections of these concepts to both aerosol physics and aerosol–cloud microphysical interactions, it is appropriate for this monograph to briefly acknowledge scientific efforts toward understanding deliberate and inadvertent modification of weather and climate (here termed “applied microphysics”), particularly since the topic has once again emerged in both scientific and public media as a means of counteracting climate change (e.g., Crutzen 2006). A more comprehensive discussion of the historical context and the evolution of weather modification research is presented in Haupt et al. (2019a).

As reviewed by Schaefer (1968), who in 1946 was the first to discover and document the nucleation of ice crystals by dry ice, the idea of using a very cold liquid to “seed” clouds had been proposed for some time; yet earlier proposals failed to understand how seeding would interact with cloud microphysics and dynamics. The discovery by Vonnegut in 1947 of the efficacy of silver iodide as an ice nucleating agent, combined with new insights into cloud and precipitation evolution, led to experimentation by these scientists and their colleagues at General Electric (GE)—who included Nobel Laureate Irving Langmuir—with the seeding of clouds (Fig. 11-35). By 1947 the cloud seeding study Project Cirrus, a joint effort between GE and the U.S. Army, Navy, and Air Force, had been launched. The project came to an end when the sudden veering of a seeded hurricane in 1947 toward the southeastern United States was blamed upon the intervention, although later it was shown that the seeding was unlikely to have had an effect on the storm track. Despite this early setback, interest in cloud seeding remained and led to a number of scientific field programs in the United States and worldwide to test hypotheses and seeding agents (e.g., Braham 1986). However, the 2003 report by the National Research Council (NRC 2003) noted that there was not conclusive evidence of a measurable effect of cloud seeding on net precipitation; it is likely that it is not possible to collect such evidence at small/regional scales. Nevertheless, active weather modification programs are currently being carried out around the world, and new studies are seeking to address the observational needs highlighted by the NRC report (Bruintjes 2016; French et al. 2018).

As also noted by Schaefer (1968), at the time of his discovery, the idea that anthropogenic activities that release large quantities of particulate matter into the atmosphere might lead to inadvertent weather modification...
was not yet fully appreciated. However, this hypothesis has been extensively explored in the decades since (Changnon 1975, 1981, 1992; NRC 2003; Warner 1968; Xue et al. 2018). A particular focus in recent years has been understanding the effects of pollution, smoke, and dust aerosols on the thermodynamic structure of the atmosphere, invigoration of convection, precipitation intensity, lightning frequency, and hurricane life cycle (e.g., Altaratz et al. 2014; Bell et al. 2009; Levin and Cotton 2009; Nowottnick et al. 2018; Posselt et al. 2012; Storer et al. 2014; Wang et al. 2009).

The concepts of direct and indirect climate forcing by aerosols described herein have their roots in these studies. Rising global concern over climate change has led to calls to consider interventions—"geoengineering"—as mitigation strategies. These include modification of incoming solar radiation by injection of aerosols in the stratosphere (Crutzen 2006; MacMartin et al. 2016; Rasch et al. 2008; Robock et al. 2008; Niemeier and Timmreck 2015; Keith et al. 2016; Visioni et al. 2017), albedo modification by changing the properties of clouds, especially marine stratus (Korhonen et al. 2010; Latham et al. 2012; Wood et al. 2017), and cirrus thinning by injecting INPs (Mitchell and Finnegar 2009; Storelmo et al. 2013; Lohmann and Gasparini 2017). The Geoengineering Model Intercomparison Project (GeoMIP; Kravitz et al. 2011) was undertaken to study, via state-of-the-science global models, the potential feedbacks and climate responses of geoengineering approaches. While many concerns regarding the environmental and societal impacts of geoengineering have been raised (e.g., Robock 2008; Russell et al. 2012), nevertheless many continue to advocate for a thoughtful and scientific approach to the study of its potential (Bellamy et al. 2012; Cicerone 2006; McNutt 2015; Vaughan and Lenton 2011; Latham et al. 2014; and community-endorsed statements by the American Meteorological Society, https://www.ametsoc.org/index.cfm/ams/about-ams/ams-statements/statements-of-the-ams-in-force/geoengineering-the-climate-system/; and the International Commission on Clouds and Precipitation, http://www.iccp-iamas.org/pdf/ICCP_RadiationManagement_Statement.pdf).

8. Outlook

Remarkably, the last century has witnessed both the birth and maturation of atmospheric chemistry and cloud physics as distinct disciplines. Tremendous progress has been realized in developing understanding of individual microphysical processes, enabled by continuing advances in laboratory methods, instrumentation, and analytical capabilities approaching the molecular scale. At the same time, the recognition of the strong coupling between aerosols, cloud physics, and dynamics, together with the availability of community airborne platforms and remarkable advancements in in situ probes, radar technologies, and satellite-borne instruments, has enabled studies of atmospheric processes that cannot be adequately addressed in the laboratory. While early attempts at forecasting weather, understanding past climate, and projecting future climates achieved initial limited success without explicitly including interactions with aerosol and cloud microphysics, the critical role that these processes play in improving the fidelity of both weather and climate forecasts is now fully recognized (e.g., Benjamin et al. 2019). Further, a full assessment of inadvertent and deliberate modification of weather and climate cannot be achieved without new advances in understanding of aerosol–cloud interactions. Thus, we anticipate that aerosol chemistry and cloud physics will continue to be strong foci of the atmospheric sciences and broader
communities into the future, as humanity seeks to address these issues with local, regional, and global implications.

Acknowledgments. We thank our Editor Greg McFarquhar, reviewers Lynn Russell and Zev Levin, and an anonymous reviewer for numerous helpful comments that greatly improved the manuscript. We also thank Susan van den Heever for significant contributions to the discussion of aerosol effects on deep convective cloud systems.

REFERENCES


Boulanger, T., and M. Stevenson, 1970: Comparison of concentrations of ice nuclei in aerosols.


Boulanger, T., and M. Stevenson, 1970: Comparison of concentrations of ice nuclei in aerosols.


Nakao, S., 2017: Why would apparent

Nagaya, U., 1954:

Myhre, G., and Coauthors, 2013: Radiative forcing of the direct

Murray, B. J., D. O'Sullivan, J. D. Atkinson, and M. E. Webb, 2012:


Xue, L., B. Geerts, X. Guo, I. Geresdi, and S. Siems, 2018: Experimental, observational, and numerical research on intentional


