Chapter 14

Radiative Forcing of Climate: The Historical Evolution of the Radiative Forcing Concept, the Forcing Agents and their Quantification, and Applications

V. RAMASWAMY, a W. COLLINS, b J. HAYWOOD, c J. LEAN, d,k N. MAHOWALD, e G. MYHRE, f V. NAIK, a K. P. SHINE, g B. SODEN, h G. STENCHIKOV, i AND T. STORELVMO j

a NOAA/Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey
b Lawrence Berkeley National Laboratory, and University of California, Berkeley, Berkeley, California
c University of Exeter, and Met Office, Exeter, United Kingdom
d U.S. Naval Research Laboratory, Washington, D.C.
e Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York
f Center for International Climate Research, Oslo, Norway
g Department of Meteorology, University of Reading, Reading, United Kingdom
h Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida
i King Abdullah University of Science and Technology, Thuwal, Jeddah, Saudi Arabia
j University of Oslo, Oslo, Norway

ABSTRACT

We describe the historical evolution of the conceptualization, formulation, quantification, application, and utilization of "radiative forcing" (RF) of Earth's climate. Basic theories of shortwave and longwave radiation were developed through the nineteenth and twentieth centuries and established the analytical framework for defining and quantifying the perturbations to Earth's radiative energy balance by natural and anthropogenic influences. The insight that Earth's climate could be radiatively forced by changes in carbon dioxide, first introduced in the nineteenth century, gained empirical support with sustained observations of the atmospheric concentrations of the gas beginning in 1957. Advances in laboratory and field measurements, theory, instrumentation, computational technology, data, and analysis of well-mixed greenhouse gases and the global climate system through the twentieth century enabled the development and formalism of RF; this allowed RF to be related to changes in global-mean surface temperature with the aid of increasingly sophisticated models. This in turn led to RF becoming firmly established as a principal concept in climate science by 1990. The linkage with surface temperature has proven to be the most important application of the RF concept, enabling a simple metric to evaluate the relative climate impacts of different agents. The late 1970s and 1980s saw accelerated developments in quantification, including the first assessment of the effect of the forcing due to the doubling of carbon dioxide on climate (the "Charney" report). The concept was subsequently extended to a wide variety of agents beyond well-mixed greenhouse gases (WMGHGs; carbon dioxide, methane, nitrous oxide, and halocarbons) to short-lived species such as ozone. The WMO and IPCC international assessments began the important sequence of periodic evaluations and quantifications of the forcings by natural (solar irradiance changes and stratospheric aerosols resulting from volcanic eruptions) and a growing set of anthropogenic agents (WMGHGs, ozone, aerosols, land surface changes, contrails). From the 1990s to the present, knowledge and scientific confidence in the radiative agents acting on the climate system have proliferated. The conceptual basis of RF has also evolved as both our understanding of the way radiative forcing drives climate change and the diversity of the forcing mechanisms have grown. This has led to the current situation where "effective radiative forcing" (ERF) is regarded as the preferred practical definition of radiative forcing in order to better capture the link between forcing and global-mean surface temperature change. The use of ERF, however, comes with its own attendant issues, including challenges in its diagnosis from climate models, its applications to small forcings, and blurring of the distinction between rapid climate adjustments (fast responses) and climate feedbacks; this
1. Radiative influences driving climate change since preindustrial times: Segue to the RF\(^2\) concept

Introduction

Interactions of the incoming solar radiation and outgoing longwave radiation with Earth's surface and atmosphere affect the planetary heat balance and therefore impact the climate system. The growth in fundamental knowledge of physics and chemistry via observational and theoretical developments through the eighteenth, nineteenth, and twentieth centuries became the platform for describing the agents driving Earth’s climate change since preindustrial times (1750) and the formulation of the “radiative forcing” (RF) (see section 2) of climate change. The central purpose of this paper is to trace the progression in the RF concept leading to our current knowledge and estimates of the major agents known to perturb climate. Below, we give a perspective into the key milestones marking advances in the knowledge of RF. Subsequent sections of the paper focus on the evolution of the concept, including its formulation, the known major forcing agents, and various applications of the concept. We attempt to capture the historical evolution of the above foci through approximately mid-2010. Of necessity, given the nature of the paper for the American Meteorological Society centennial monograph volume and the vast domain of the topic, the principal aim of this manuscript is to describe the evolution as evidenced through the literature, particularly the major international assessment reports. We refer the reader to the richness of the references cited for the in-depth scientific details marking the steps over the past three centuries to the present state of the art.

1) Growth of Atmospheric Radiation Transfer (Pre-Twentieth Century to Mid-Twentieth Century)

The basic concepts of planetary energy budget and the greenhouse effect were put forward in the early nineteenth century by Fourier (1824), although the term “greenhouse” was not mentioned. Fourier recognized that the atmosphere is opaque to “dark heat” (infrared radiation) but could not identify the factors. Laboratory experiments related to transmission of light by atmospheric gases at different wavelengths were the subject of atmospheric radiation inquiries as far back as the early nineteenth century. One of the very first laboratory measurements of infrared absorption was reported by Tyndall (1861). Based on a series of carefully designed laboratory experiments, Tyndall discovered that infrared absorption in the atmosphere is largely due to carbon dioxide and water vapor. Tyndall thought that variations in the atmospheric concentrations of CO\(_2\) and water vapor account for “all the mutations of climate which the researches of geologists reveal” (see Anderson et al. 2016). Very soon after that came spectral measurements, prompted by both scientific curiosity and a quest to explain the then known variations in Earth’s climate.

Arrhenius (1896) made the quantitative connection to estimate the surface temperature increase due to increases in CO\(_2\). He relied on surface radiometric observations (Langley 1884), used or inferred a number of fundamental principles in shortwave and longwave radiations. The forcing-response linkage as simulated by models, including the diversity in the spatial distribution of forcings by the different agents, has provided a practical demonstration of the effectiveness of agents in perturbing the radiative energy balance and causing climate changes. The significant advances over the past half century have established, with very high confidence, that the global-mean ERF due to human activity since preindustrial times is positive (the 2013 IPCC assessment gives a best estimate of 2.3 W m\(^{-2}\), with a range from 1.1 to 3.3 W m\(^{-2}\); 90% confidence interval). Further, except in the immediate aftermath of climatically significant volcanic eruptions, the net anthropogenic forcing dominates over natural radiative forcing mechanisms. Nevertheless, the substantial remaining uncertainty in the net anthropogenic ERF leads to large uncertainties in estimates of climate sensitivity from observations and in predicting future climate impacts. The uncertainty in the ERF arises principally from the incorporation of the rapid climate adjustments in the formulation, the well-recognized difficulties in characterizing the preindustrial state of the atmosphere, and the incomplete knowledge of the interactions of aerosols with clouds. This uncertainty impairs the quantitative evaluation of climate adaptation and mitigation pathways in the future. A grand challenge in Earth system science lies in continuing to sustain the relatively simple essence of the radiative forcing concept in a form similar to that originally devised, and at the same time improving the quantification of the forcing. This, in turn, demands an accurate, yet increasingly complex and comprehensive, accounting of the relevant processes in the climate system.
radiation, pointed out the greenhouse effect of water vapor and CO₂, and made simple assumptions concerning exchange of heat between surface and atmosphere to deduce the temperature change (see Ramanathan and Vogelmann 1997). In the same study, Arrhenius also discussed the solar absorption in the atmosphere. Arrhenius’ systematic investigation and inferences have proven to be pivotal in shaping the modern-day thinking and computational modeling of the climate effects due to CO₂ radiative forcing.

Advances in theoretical developments in classical and, later on, in quantum physics through the nineteenth and early twentieth centuries laid the groundwork concerning light (photon) absorption/emission processes and their linkage to the laws of thermodynamics. This led to the enunciation of basic concepts in the nineteenth century, for example, Kirchhoff’s deductions concerning blackbody radiation, and the associated laws by Planck, Wien, Rayleigh–Jeans, and Stefan–Boltzmann. These laws, and the physics of thermal absorption and emission by gases and molecules, were applied to the context of the atmosphere, leading to the formalism of atmospheric longwave radiative transfer [see chapter 2 in Goody and Yung (1995)].

Discovery and understanding of observed phenomena played a role throughout in the development of methodologies that were to become building blocks for the quantification of perturbations to the shortwave and longwave radiative fluxes. A combination of fundamental theoretical developments, observations, simple calculations, and arguments sowed the advances, for example, Lord Rayleigh’s (Hon. J. W. Strutt) treatise on sky light and color (Rayleigh 1871) and the electromagnetic scattering of light (Rayleigh 1881). Another example is Mie’s theory of electromagnetic extinction (Mie 1908), which unified the laws of light reflection, refraction, and diffraction following Huygens, Fresnel, and Snell (see van de Hulst 1957) and inferred the disposition of light at any wavelength when it interacts with homogeneous spherical particles. Advances in the knowledge of gaseous absorption and emission processes through laboratory-based quantification of absorption lines and band absorption by the important greenhouse gases marked the further growth of atmospheric longwave radiative transfer from the late nineteenth century into the mid- and late twentieth century [see chapter 3 in Goody and Yung (1995)]. Experimental developments, along with advances in conceptual thinking on the heat balance of the planet, began to provide the platform for quantifying the radiation budget, for example, solar irradiance determination by Abbott and Fowle (1908), and an early estimate of Earth’s global-average energy budget by Dines (1917). Dines’s effort was a remarkable intellectual attempt given there was very little then by way of observations of the individual components. Figure 14-1 provides a comparison of the values estimated by Dines (1917) compared to one modern

FIG. 14-1. Comparison of one early estimate of Earth’s global-average energy budget (Dines 1917) with the contemporary estimates of L’Ecuyer et al. (2015) by annotating the original figure from Dines (1917). All values are given in W m⁻², with Dines’s values in plain font, and L’Ecuyer et al. in bold font. Dines’s value for the surface LW emission is low probably because he adopted a value for Stefan’s constant that was “decidedly lower than that usually given,” although the assumed surface temperature is not stated either. For some components, Dines also gave an estimate of the uncertainty. The L’Ecuyer et al. (2015) values are from their Fig. 4, which applies energy and water balance constraints. [From Dines (1917), copyright Royal Meteorological Society.]
analysis (L’Ecuyer et al. 2015). What we term as RF of climate change today can be regarded as a result of this early thinking about the surface–atmosphere heat balance.

Callendar’s work in the 1930s–50s (Fleming 2007) built upon the earlier explorations of Arrhenius (1896) and Ekholm (1901) to relate global temperature to rising CO₂ concentrations. Callendar (1938) compiled measurements of temperatures from the nineteenth century onward and correlated these with measurements of atmospheric CO₂ concentrations. He concluded that the global land temperatures had increased and proposed that this increase could be an effect of the increase in CO₂ (Fleming 1998). Callendar’s assessment of the climate sensitivity (defined as surface temperature change for a doubling of CO₂) was around 2°C (Archer and Rahmstorf 2010), which is nowadays regarded as being at the lower end of the modern-day computed values [e.g., Intergovernmental Panel on Climate Change (IPCC); IPCC 2013]. His papers in the 1940s and 1950s influenced the study of CO₂–atmosphere–surface interactions vigorously, both on the computational side, which introduced simplified radiation expressions (e.g., Plass 1956c; Yamamoto and Sasamori 1958), and in initiating the organization of research programs to measure CO₂ concentrations in the atmosphere. Plass recognized the importance of CO₂ as a greenhouse gas in 1953 and published a series of papers (e.g., Plass 1956a,b,d). He calculated that the 15-micron CO₂ absorption causes the temperature to increase by 3.6°C if the atmospheric CO₂ concentration is doubled and decreases by 3.8°C when it is halved. These early calculations helped guide future works. Modern monitoring of CO₂ concentrations began with Keeling’s pioneering measurements of atmospheric CO₂ concentrations, begun in connection with the International Geophysical Year in 1957 (e.g., Keeling 1960). This soon spurred the modern computations of the effects due to human-influenced CO₂ increases and initiated investigations into anthropogenic global warming. The historical developments above, plus many others, beginning principally as scientific curiosity questions concerning Earth’s climate, have formed the foundational basis for the contemporary concept of RF and the estimation of the anthropogenic effects on climate.

A major part of the work related to radiative drivers of climate change came initially on the longwave side, and more particularly with interest growing in the infrared absorption by CO₂ and H₂O. This came about through the works of many scientists [see references in chapters 3 and 4 of Goody and Yung (1995)]. Research expansion comprising theoretical and laboratory measurements continued into the late twentieth century [see references in chapter 5 of Goody and Yung (1995)]. Importantly, from the 1960s, existing knowledge of spectral properties of gaseous absorbers began to be catalogued on regularly updated databases, notably high-resolution transmission molecular absorption (HITRAN) (see references in chapter 5 of Goody and Yung 1995).

On the shortwave measurements side, the Astrophysical Observatory of the Smithsonian Institution (APO) made measurements of the solar constant (now more correctly referred to as the “total solar irradiance” as it is established that this is not a constant) at many locations on Earth’s surface from 1902 to 1962 (Hoyt 1979). While there were interpretations from these observations about change and variations in the sun’s brightness, the broad conclusion was that the data reflected a strong dependency on atmospheric parameters such as stratospheric aerosols from volcanic eruptions, as well as dust and water vapor. Research into shortwave and longwave radiation transfer yielded increasingly accurate treatments of the interactions with atmospheric constituents ( chapters 4–8 in Goody and Yung 1995; chapters 1–4 in Liou 2002).

2) ADVENT OF THE RF CONCEPT AND ITS EVOLUTION (SINCE THE 1950S)

Advances in computational sciences and technology played a major role alongside the growth in basic knowledge. The increases in computational power from the 1950s onward, with facilitation of scalar and later vector calculations, enhanced the framework of “reference” computations (e.g., Fels et al. 1991; Clough et al. 1992). This enabled setting benchmarks for quantifying the radiative forcing by agents. With developments in community-wide radiative model intercomparisons (e.g., Ellingson et al. 1991; Fouquart et al. 1991; Collins et al. 2006), the comparisons against benchmarks established a definitive means to evaluate radiative biases in global weather and climate models, one of the best examples of a “benchmark” and its application in the atmospheric sciences. The advance in high-performance computing since 2000 has endowed the benchmark radiative computations with the ability to capture the details of molecular absorption and particulate extinction at unprecedented spectral resolutions in both the solar and longwave spectrum.

Relative to the previous decades, the 1950s also witnessed the beginning of increasingly sophisticated and practical numerical models of the atmosphere and surface that included radiative and then radiative–convective equilibrium solutions. Hergesell’s (1919) work had superseded earlier calculations in describing the radiative equilibrium solutions using a gray-atmosphere approach. Subsequent studies further advanced the field by recognizing the existence of a thermal structure, making more realistic calculations based on newer
spectroscopic measurements and observations (e.g., Murgatroyd 1960; Mastenbrook 1963; Telegadas and London 1954), developing simplified equations (parameterizations) for use in weather and climate models, and exploring how the radiation balance could be perturbed through changes in the important atmospheric constituents, for example, water vapor and carbon dioxide (Kaplan 1960; Kondratiev and Niilisk 1960; Manabe and Møller 1961; Houghton 1963; Møller 1963; Manabe and Strickler 1964). Manabe and Strickler (1964) and Manabe and Wetherald (1967) set up the basis for the more modern-day calculations in the context of one-dimensional models, invoking radiative–convective equilibrium, where the essential heat balance in the atmosphere–surface system involved solar and longwave radiative and parameterized convective (latent + sensible heat) processes. In this sense, the 1960s efforts went significantly ahead of Arrhenius’s pioneering study and other earlier insightful investigations to recognize and calculate the effects of carbon dioxide in maintaining the present-day climate.

The foundational model calculations of radiative perturbations of the climate system arose from publications beginning in the 1960s. Manabe and Wetherald (1967) demonstrated how changes in radiative constituents (CO₂, H₂O, O₃) as well as other influences (solar changes, surface albedo changes) could affect atmospheric and surface temperatures. The field of modeling grew rapidly over the 1960s to 1980s period and three-dimensional models of the global climate system came into existence, enabling an understanding of the complete latitude–longitude–altitude effects of increasing CO₂. The acceleration of modeling studies resulted in an ever-increasing appreciation of CO₂ as a major perturbing agent of the global climate (Manabe and Bryan 1969; Manabe and Wetherald 1975; Ramanathan et al. 1979; Manabe and Stouffer 1980; Hansen et al. 1981; Bryan et al. 1988; Washington and Meehl 1989; Stouffer et al. 1989; Mitchell et al. 1990). The growth in the number of studies also galvanized CO₂-climate assessments using the numerical model simulations (e.g., NRC 1979; now famously referred to as the “Charney” report). The Charney study was the first institutionally sponsored scientific assessment based on then available studies. The report concluded an RF due to CO₂ doubling of about 4 W m⁻² and estimated the most probable global warming to be near 3°C with a probable error of ±1.5°C. This was a landmark report, has influenced the community immensely, and became a trendsetter for climate science assessments. A second assessment followed (NRC 1982; referred to as the “Smagorinsky” report) that essentially reiterated the conclusions of the Charney report.

The above studies and assessments established a useful basis for a formalized perspective into mathematical linkages between global-mean RF by greenhouse gases and surface temperature changes, with the applicability extending to global climate impacts. The modern definition and equations for RF took root during this period. The conceptual development that has lent powerful significance to characterizing radiative perturbations via “RF” came through in the 1970s with the first formal phrasing (Ramanathan 1975), and got solidified as a concept in the late 1970s and 1980s (e.g., Ramanathan et al. 1979; Dickinson and Cicerone 1986), especially through the major international assessment reports (e.g., Ramanathan et al. 1985a,b). Eventually, the IPCC scientific assessments, beginning with IPCC (1990), made this a robust terminology.

This continues through today even though there have been substantial refinements in the past decade (see section 2). As the RF concept settled into more rigorous formulations in the 1970s and 1980s, a spate of research extended this exercise to other well-mixed greenhouse gas changes such as methane, nitrous oxide, and chlorofluorocarbons (Ramanathan 1975; Wang et al. 1976; Donner and Ramanathan 1980; Hansen et al. 1981). This became possible as spectroscopic data and knowledge of their atmospheric concentration changes grew with time. In later years and decades, the list of well-mixed greenhouse gases grew to include a plethora of halocarbons, sulfur hexafluoride, etc. (e.g., Fisher et al. 1990; Pinnock et al. 1995).

Although the RF concept was developed to quantify the changes in radiation balance due to well-mixed greenhouse gases and solar irradiance changes, this was extended to short-lived gases, such as ozone, which exhibit strong spatial and temporal variability (Ramanathan et al. 1985b; Shine et al. 1990; Isaksen et al. 1992). The concept was also applied to an entire category of effects referred to as “indirect” that accounted for changes in atmospheric concentrations of a radiative constituent affected by non-radiative effects such as chemical or microphysical interactions (see sections 4 and 5). These were first derived for the case of tropospheric and stratospheric ozone changes occurring through chemical reactions in the atmosphere involving anthropogenic precursor species. Indirect effects also were uncovered for aerosol-related radiative effects obtained through their interactions with water and ice clouds (Charlson et al. 1992; Penner et al. 1992; Schimel et al. 1996).

The impact of emissions of anthropogenic aerosols, or their precursors, on climate had been recognized as early as the 1970s, while recognition of their effects on air pollution goes back more than a century (Brimblecombe and Bowler 1990). The first quantification, however, in
the context of preindustrial to present-day emissions came through Charlson et al. (1991). The forcing connected with the anthropogenic aerosol emissions has acquired a more diverse picture now with the complexity associated with the various species (e.g., different types of carbonaceous aerosols), existence of a variety of mixed states (i.e., aerosols consisting of more than one component), and the influence of each species on the formation of water drops and ice crystals (“indirect” forcing referred to above). Additional complexities with aerosols as compared to the well-mixed greenhouse gases arise because of their inhomogeneous space and time distribution. Estimating preindustrial concentrations of important short-lived gases and aerosols and their precursors is difficult and is a major contributor to uncertainty in their RF (e.g., Tarasick et al. 2019; Carslaw et al. 2017).

Besides atmospheric constituents, other radiative influences also began to be quantified under the broad concept of “radiative forcing.” These included land-use and land-cover changes due to vegetation changes, primarily in the Northern Hemisphere. The initial considerations were for the changes induced in the albedo of the surfaces due to human activity (Sagan et al. 1979). Later, other physical factors in the context of forced changes such as surface roughness, trace gas and aerosol emissions, water, and water-related changes as a consequence of land surface changes were also considered as it was realized that these too affected the planetary heat balance (e.g., IPCC 2013).

A relatively recent entry under the anthropogenic RF label includes the attempts to quantify the forcing due to aviation-induced aerosols and contrails, reported as early as beginning of 1970s, and quantitatively assessed beginning with IPCC (1999) (e.g., Fahey et al. 1999). Emissions from various industrial sectors including transportation (aircraft, shipping, road transport) have been comparatively evaluated and assessed (see Unger et al. 2010). While anthropogenic forcings became increasingly better quantified in the twentieth century, so too were the natural agents, such as solar irradiance changes (see Hoyt and Schatten 1997) and aerosols formed in the stratosphere in the aftermath of explosive or climatically significant volcanic eruptions (Franklin 1784; Robock 2000). The qualitative recognition of the potential climatic effects due to powerful volcanic eruptions (e.g., Toba, Tambora, and Krakatoa eruptions), and solar changes, possibly goes more than two centuries back. As an example, solar irradiance changes and the resultant transmission of sunlight through the atmosphere began to be pursued as both questions of scientific curiosity and for potential impacts on surface climate.

3) SCOPE OF THE PAPER

In this paper, we trace the evolution of the knowledge base that began with recognizing the importance of changes in atmospheric composition, how they alter the radiative balance of the planet, and the resulting growth in understanding that has enabled quantification of the radiative effects. Notably, this began with considerations of the roles of water vapor and carbon dioxide in the longwave spectrum and the naturally arising solar irradiance changes and particulates from volcanic eruptions in the shortwave spectrum. The early discoveries and theories on the role of radiation in the planet’s heat equilibrium state paved the way for defining the forcing of Earth’s climate system, with gradually increasing attention to the range of anthropogenic influences. The forcing used in this context was meant to characterize the agents driving climate change and nominally on a global-average basis, rather than regional or local scales. In describing the evolution of the RF concept and its applications, we follow a strategy of describing the principal advancements over time, with references to a few of the seminal investigations. Included in these are the well-known chapters on radiative forcing appearing in various assessments and reports (e.g., IPCC 1990, 1992, 1994, 1996, 1999, 2001, 2007, 2013; Ramanathan et al. 1985b; NRC 1979). Our aim is not to summarize from the assessments but instead to document the key elements happening over time that pushed the frontiers to the state of the art in its successive evolutionary stages through to today. We hew fairly strictly to RF only. We do not discuss “climate feedbacks” per se, which are an integral part of climate response, but that discussion is outside the scope of this paper.

Figure 14-2 illustrates the radiative forcing quantification in each of the five major IPCC WGI assessments to date (1990, 1996, 2001, 2007, and 2013). All the forcings on the illustration represent a measure of the radiative perturbation at the tropopause brought about by the change in that agent relative to its value/state in 1750. As the knowledge has advanced, there has been a growth in the number of forcing agents and an evolution in the estimates of the magnitudes of the agents. The increased attention to scientific uncertainties also becomes evident, representing an advance in the measure of scientific understanding. Quantification of the anthropogenic WMGHG and the secular solar forcing began from the first IPCC assessment [IPCC 1990; or First Assessment Report (FAR)]. While aerosol radiative effects were recognized in FAR, the tropospheric aerosol quantification was reported in a supplementary IPCC Report (IPCC 1992), then in a special report.
(IPCC 1994), which was reaffirmed in the Second Assessment Report (SAR; IPCC 1996). RF from ozone changes was recognized in FAR but quantified later. The RF from stratospheric ozone losses due to the halocarbon-catalyzed chemical reactions and that due to tropospheric ozone increases from anthropogenic precursor emission increases and related chemistry–climate interactions was first quantified in IPCC (1992) followed by IPCC (1994). A special report on aviation-related impacts appeared as IPCC (1999).

The Third Assessment Report (TAR; IPCC 2001) added a few more agents that were able to be quantified besides updating the estimates of the greenhouse gas and aerosol agents. This occurred in part due to accounting for the increased knowledge about changes in the species concentrations and, to a lesser extent, due to improvements in the treatment of the processes.

The Fourth Assessment Report (AR4; IPCC 2007) introduced new methodologies to estimate short-lived gas RF and to express the uncertainty due to tropospheric aerosols, which continue to be the principal reason for the large uncertainty in the anthropogenic forcing (section 10). The Fifth Assessment Report (AR5; IPCC 2013) introduced a major change in the manner of expressing the radiative forcing by making the transition from RF to the effective radiative forcing (ERF). Further details on the progress through the IPCC assessments appear in section 2. The change in radiative forcing due to CO2 is due to increase in the concentration between the IPCC assessments, except between SAR and TAR, where there was an update in the expression for calculating the radiative forcing. On the other hand, the changes in the short-lived compounds such as ozone and aerosols from one assessment to the other are mainly results of improvements based on observations and modeling representing the knowledge prevailing at the time of the IPCC assessments.

The presentation in this paper aims to capture the principal developments of each forcing and their chief characteristics as they developed over time and thus does not insist on discussions of all forcing agents to hew to the same presentation format in the discussions. The sections below discuss the major facets of the radiative forcing concept, beginning with its formulation in section 2. Sections 3, 4, 5, 6, and 7 address the development of the quantified knowledge, including uncertainties, of the anthropogenic forcing agents in tandem with the developments in the IPCC assessments beginning with the First Assessment Report in 1990. Sections 8 and 9 discuss the natural drivers of climate change. The totality of the forcing of the climate system, that is, a synthesis by accounting for all the agents in a scientifically justified manner, is examined in section 10. The role of RF in enabling the development of metrics to allow emissions of different gases to be placed on an equivalent scale is discussed in section 11, while the connection of response to the forcing culled from observations and climate model simulations follows in section 12. Section 13 traces the development of a relatively new concept in the application of forcing, namely, management of solar and terrestrial radiation based on the forcing regarding well-mixed greenhouse gases and aerosols. The concluding section summarizes the major points of the paper: development, utilization, and application of radiative forcing; strengths and limitations of the forcing concept; unresolved issues and grand challenge related to the viability of this concept; and its use for the quantification of climate change.

2. Radiative forcing—Its origin, evolution, and formulation

a. The utility of the forcing-feedback-response framework

Radiative forcing provides a metric for quantifying how anthropogenic activities and natural factors perturb the flow of energy into and out of the climate system. This perturbation initiates all other changes of the climate due to an external forcing. The climate system responds to restore radiative equilibrium through a change in temperature, known as the Planck response or Planck feedback. A positive forcing (i.e., a net radiative gain) warms the climate and increases the thermal emission to space until a balance is restored. Similarly, a negative forcing (i.e., a net radiative loss) cools the climate, decreasing the thermal emission until equilibrium is restored.

The change in temperature required to restore equilibrium can induce other surface and atmospheric changes that impact the net flow of energy into the climate system, and thus modulate the efficiency at which the climate restores equilibrium. Borrowing terminology from linear control theory, these secondary changes can be thought of as feedbacks that serve to further amplify or dampen the initial radiative perturbation. The use of radiative forcings and radiative feedbacks to quantify and understand the response of climate to external drivers has a long and rich history (Schneider and Dickinson 1974; Hansen et al. 1984; Cess 1976; Cess et al. 1990; NRC 2005; Stephens 2005; Sherwood et al. 2015).

Consider a perturbation in the global-mean net downward irradiance at the top of atmosphere, \( dF \) (which we call the RF), that requires a change in global-mean surface temperature, \( dT \), to restore radiative equilibrium (overbars indicate a global-average quantity). If
FIG. 14-2. Summary of the evolution of the global-mean radiative forcings from IPCC reports, where available, from FAR (1765–1990), SAR (1750–1992), TAR (1750–1998), AR4 (1750–2005), and AR5 (1750–2011). The RF and/or the ERF presented in AR5 are included. Uncertainty bars show the 5%–95% confidence ranges. (a) From top to bottom, the forcings are due to changes in CO₂, non-CO₂ WMGHGs, tropospheric ozone, stratospheric ozone, aerosol–radiation interaction, aerosol–cloud interaction, surface albedo, total anthropogenic RF, and solar irradiance. The forcings are color coded to indicate the “confidence level” (or LOSU as was presented in and before AR4, which used “consensus” rather than “agreement” to assess confidence level). Dark green is “high agreement and robust evidence”; light green is either “high agreement and medium evidence” or “medium agreement and robust evidence”; yellow is either “high agreement and limited evidence,” “medium agreement and medium evidence,” or “low agreement and robust evidence”; orange is either “medium agreement and limited evidence” or “low agreement and medium evidence”; and red is “low agreement and limited evidence.” Several minor forcings (such as due to contrails and stratospheric water vapor due to methane changes) are not included. The information used here, and information on excluded components, can be mostly found in Myhre et al. (2013, their Tables 8.5 and 8.6 and Fig. 8.14) and Shine et al. (1990, their Table 2.6). The decrease in CO₂ RF between SAR and TAR was due to a change in the simplified expressions used to compute its RF; the CO₂ concentration has increased monotonically between each successive IPCC report. No central estimate was provided for aerosol–cloud interaction in SAR and TAR, and a total aerosol–radiation interaction [see (b)] and a total anthropogenic RF was not presented in assessments prior to AR4. Stratospheric aerosol RF resulting from volcanic aerosols →
the changes are small and higher-order terms can be neglected, and \( d\mathcal{F} \) is time independent, the change in upward radiative energy, \( d\mathcal{R} \), induced by the change in surface temperature, \( dT \), can be decomposed into linear contributions from changes in temperature and other radiative feedbacks \( X_i \):

\[
d\mathcal{R} = \frac{\partial \mathcal{R}}{\partial T} dT + \sum_i \frac{\partial \mathcal{R}}{\partial X_i} \frac{\partial X_i}{\partial T} dT.
\]  

Equilibrium is restored when \( d\mathcal{R} = d\mathcal{F} \). The ratio, \( \alpha = d\mathcal{R}/dT \), called the “climate feedback parameter,” quantifies the efficiency at which the climate restores radiative equilibrium following a perturbation. In the absence of feedbacks, the Planck response is \( \alpha_i \approx 3.3 \text{ W m}^{-2}\text{K}^{-1} \) (e.g., Cess 1976). In current climate models, radiative feedbacks from water vapor, clouds, and snow/sea ice cover act to reduce \( \alpha \) to a range \( \approx 1-2 \text{ W m}^{-2}\text{K}^{-1} \); this amplifies the change in temperature in response to a given radiative forcing. Most of the intermodel spread in \( \alpha \) is due to differences in predicting the response of clouds to an external forcing (Cess et al. 1990). Feedbacks from water vapor, clouds, snow, and sea ice cover have been well documented in both models (Bony et al. 2006) and, to a lesser extent, in observations (Forster 2016). Less well studied are feedbacks from the carbon cycle, ice sheets, and the deep ocean that occur on much longer time scales (e.g., Gregory et al. 2009; Forster 2016).

While attempting to characterize global climate changes using a single scalar quantity may seem overly simplistic, many aspects of climate do respond in proportion to \( dT \), regardless of the spatial and temporal scales being considered and are of much greater societal relevance than global-mean temperature (e.g., the magnitude of

---

\("\) is not included due to their episodic nature; estimates can be seen in Fig. 14-31. \( b \) Individual components of RF due to changes in aerosol–radiation interaction. From top to bottom, these are sulfate, black carbon from fossil fuel or biofuel burning, biomass burning, organic aerosols, dust, nitrate, and total [also shown in (a)]. In AR5, the organic aerosol RF was separated into primary organic aerosol (POA) from fossil fuel and biofuel and secondary organic aerosol (SOA), due to changes in source strength, partitioning, and oxidation rates. Separate confidence levels were not presented for individual components of the aerosol–radiation interaction in AR4 and AR5, and hence, none are shown. The information used here is mostly drawn from Myhre et al. (2013, their Table 8.4).
regional rainfall change). To the extent that \(dF\) can be used to estimate \(dT\), radiative forcing then provides a simple but crude metric for assessing the climate impacts of different forcing agents across a range of emission scenarios. Here, we write the relationship between \(dT\) and \(dF\) as

\[
dT = \lambda dF, \quad (14-2)
\]

where \(\lambda\) is usually referred to as the “climate sensitivity parameter,” the inverse of \(\alpha\). It is worth noting that equilibrium climate sensitivity is often written in terms of the equilibrium surface temperature response, in K, to a doubling of CO_2 (about 3.7 W m\(^{-2}\)).

An important driver in the early development of RF as a metric was the chronic uncertainty in the value of \(\lambda\), which persists to this day; this meant that quantifying the drivers of climate change, and intercomparing different studies, was easier using \(dF\) rather than \(dT\). However, such a comparison of different climate change mechanisms relies on the extent to which \(\lambda\) is invariant (in any given model) to the mechanism causing the forcing; early studies demonstrated similarity between the climate sensitivity parameter for CO_2 and solar forcing (e.g., Manabe and Wetherald 1975), but subsequent work [see section 2c(4)] has indicated limitations to this assumption.

The conceptual development in the subject, which will be discussed in the following sections, has adopted progressively more advanced definitions of RF with the aim of improving the level of approximation in Eq. (14-2).

b. Origin of the radiative forcing concept (1970s–80s)

Ramanathan (1975) presents the first explicit usage of the RF concept, as currently recognized (although the term “radiative forcing” was not used), in an important paper quantifying, for the first time, the potential climate impact of chlorofluorocarbons (CFCs). Ramanathan computed the change in the top-of-atmosphere (TOA) irradiance due to increased CFC concentrations and directly related this to the surface temperature change, via an empirical estimate of the dependence of the irradiance on surface temperature; this is the climate feedback parameter discussed in section 2a. Ramanathan noted that the surface temperature calculations using this “simpler procedure” were identical to those derived using a “detailed” radiative–convective model. Ramanathan and Dickinson (1979) extended the Ramanathan (1975) framework in important ways in a study of the climate impact of stratospheric ozone changes. First, there was an explicit recognition that changes in stratospheric temperature (in this case, driven by stratospheric ozone change) would influence the tropospheric energy balance. Second, these calculations were latitudinally resolved. While the global-average stratosphere is in radiative equilibrium (and hence temperature changes can be estimated via radiative calculations alone), locally dynamical heat fluxes can be important. Ramanathan and Dickinson considered two “extreme” scenarios to compute this temperature change without invoking a dynamical model. One assumed that dynamical feedbacks were so efficient that they maintained observed latitudinal temperature gradients; given subsequent developments, this is of less interest here. The other scenario assumed that, following a perturbation, dynamical fluxes remain constant, and temperatures adjust so that the perturbed radiative heating rates equal unperturbed heating rates (and thus balance the unperturbed dynamical heat fluxes). This second method was originally referred to as the “no feedback case” (the “feedback” referring to the response of stratospheric dynamics to a forcing); it has since become more widely known as “fixed dynamical heating” (FDH) (Fels et al. 1980; WMO 1982; Ramanathan et al. 1985b or more generally “stratospheric (temperature) adjustment.” FDH has also been used for stratospheric temperature trend calculations and shown to yield reasonable estimates of temperature changes derived from a general circulation model (GCM) (e.g., Fels et al. 1980; Kiehl and Boville 1988; Chanin et al. 1998; Maycock et al. 2013).

Ramanathan et al. (1979) applied the same methodology to CO_2 forcing. Their estimate of RF for a doubling of CO_2 of about 4 W m\(^{-2}\) was adopted in the influential NRC (1979) report and has been an important yardstick since then. Although not explicitly stated until subsequent papers in the 1980s (see later), one key reason for including stratospheric temperature adjustment as part of RF, rather than as a climate feedback process, was that the adjustment time scale is of order months; this is much faster than the decadal or longer time scale for the surface temperature to respond to radiative perturbation, which is mostly driven by the thermal inertia of the ocean mixed layer. A second, related, key reason is that the tight coupling of the surface and troposphere, via convective heat fluxes (and, conversely, the limited coupling between the surface and the stratosphere), means that \(dT\) at surface is largely driven by the RF at the tropopause. A consequence of applying stratospheric temperature adjustment (which returns the stratosphere to global radiative equilibrium) is that tropopause and top-of-atmosphere forcings are identical. This removes an important ambiguity in the definition of RF, although the definition of the tropopause still has to be considered [see section 2c(5)].

These time scales were made explicit by Hansen et al. (1981), who demonstrated the evolution of the irradiance
changes in their radiative–convective model, following a doubling of CO₂ (see Fig. 14-3). This tracked the changes from (i) the immediate response [nowadays called the instantaneous RF (IRF)]; (ii) the response after a few months (which is close to the RF incorporating stratospheric temperature adjustment); in the case of a CO₂ increase, the increased emittance of the stratosphere leads to a cooling that increases the magnitude of the perturbation of the top-of-atmosphere irradiance from \(2.4 \text{ W m}^{-2}\); and (iii) “many years later” when the surface temperature has equilibrated [following Eq. (14-2)] and the resulting irradiance change at the top of the atmosphere has cancelled out the forcing. Hansen et al. (1981) seem to be the first to use the terminology “radiative forcing,” although they used it in a general rather than a quantitative sense.

Several contributions to the edited volume by Clark (1982) also used the RF concept, at least in an illustrative way, although using a variety of names. For example, Chamberlain et al. (1982) compared different climate change mechanisms using what would now be called as surface radiative forcing; their use of this (rather than tropopause or top-of-atmosphere RF) as a predictor of surface temperature change was strongly disputed in the same volume by Ramanathan (1982) (and earlier in Ramanathan 1981), Hansen et al. (1982), and Luther (1982). Hansen et al. (1982) briefly presented values of top-of-atmosphere radiative flux changes for idealized changes in concentrations of five gases and refer to these as “radiative forcings” in their text. At around the same time, WMO (1982), in a brief report on a meeting on the potential climate effects of ozone and other minor trace gases, refer explicitly to “the net outgoing longwave flux at the tropopause” as determining “the radiative forcing of the surface–troposphere system” and present values for idealized perturbations of six greenhouse gases.

Ramanathan et al. (1985a) also used the term “radiative forcing” in the context of Eq. (14-2), and Dickinson and Cicerone (1986) appear to be the first to use the concept to quantify the climate impact of changes in concentrations of several greenhouse gases relative to preindustrial times in watts per square meter (using the term “trapping” rather than RF).

“Radiative forcing” became firmly established as accepted terminology in chapter 15 of the 1985 WMO Ozone Assessment (Ramanathan et al. 1985b) [which was largely reproduced in Ramanathan et al. (1987)], and the term was widely used in their discussion; however, much of their overall comparison of the impacts of climate forcing agents was still posed in terms of surface temperature change.

c. The evolution of the radiative forcing concept during the IPCC era

Assessment of RF has been firmly embedded in IPCC assessments from its FAR onward. FAR (Shine et al. 1990) took as its starting point the fact that the climate impact of a range of different climate forcing agents could be compared using RF, in watts per square meter, even though this was only starting to be done routinely in the wider literature at the time. A significant motivator for the use of RF in all IPCC reports was as input to climate emissions metrics (such as the global warming potential; see section 11). This section focuses mostly on developments in understanding of anthropogenic forcings—more detailed discussions of the evolution of the understanding of solar and volcanic aerosol forcings is given in sections 8 and 9, respectively.

1) The IPCC FAR

FAR discussed the concept of RF, stressing the utility of including stratospheric temperature adjustment. Building on earlier work (e.g., Ramanathan et al. 1987), it also emphasized the importance of indirect forcings, such as the impact of changes in methane concentrations on ozone and stratospheric water vapor. The main focus was on

---

FIG. 14-3. Changes in top-of-atmosphere shortwave (\(\Delta S\)) and longwave (\(\Delta F^l\)) irradiances following a doubling of CO₂ from 300 ppm, in a one-dimensional radiative–convective model. Wavy lines represent changes in convective fluxes, with all other lines radiative. The values in boxes show the net flux changes at the surface and for the atmosphere. (a) The instantaneous flux change; (b) the change after a few months, including the effect of stratospheric temperature adjustment and other rapid adjustments represented in the model; and (c) the flux changes when the system has come to equilibrium following a change in surface temperature (\(T_S\)). [Figure taken from Hansen et al. (1981); reprinted with permission from AAAS.]
greenhouse gases, including extended tabulations of forcing due to CFCs and their potential replacements. FAR also popularized the use of simplified expressions for calculating RF, which were empirical fits to more complex model calculations. The expressions used in FAR were based on two studies available at the time (Wigley 1987; Hansen et al. 1988). Updated versions of the simplified expressions are still widely used in simple climate models and for assessing potential future scenarios of trace gas concentrations.

FAR also included, together in a single section, the roles of solar variability, direct aerosol effects, indirect aerosol effects, and changes in surface characteristics. The literature on these was sparse. The section on direct aerosol forcing focused mostly on volcanic aerosol; it did not attempt to quantify the impact of human activity because of “uncertainties in the sign, the affected area and the temporal trend.” Perhaps surprisingly, more attention was given to the indirect aerosol effects (now more generally known as “aerosol–cloud interactions”); although FAR stated that “a confident assessment cannot be made,” due to important gaps in understanding, a 1900–85 estimate of $-0.25$ to $-1.25$ W m$^{-2}$ (based on Wigley 1989) was provided. FAR did not include estimates for preindustrial to present-day RF across all forcing agents but restricted itself to two then-future periods (1990–2000 and 2000–50).

Soon after FAR, an IPCC supplementary report provided an update (Isaksen et al. 1992). Significant developments since FAR included more advanced RF estimates due to ozone change (Lacis et al. 1990) and the first calculation of the forcing due to latitudinally resolved observed stratospheric ozone depletion (Ramawat et al. 1992). The indirect forcing due to methane’s impact on tropospheric ozone and stratospheric water vapor were quantified. The first geographically resolved estimates of sulfate aerosol direct forcing (now referred to as aerosol–radiation interaction) (Charlson et al. 1991) had become available, indicating a significant offset (in the global-mean sense) of greenhouse gas RF.

2) IPCC SPECIAL REPORT ON RADIATIVE FORCING AND THE IPCC SAR

The SAR discussion on RF was partly based on the analysis of Shine et al. (1994) in an IPCC special report. Since FAR, there had been several important developments. The 1992 Pinatubo volcanic eruption had allowed unprecedented global-scale observations of the impact of such a large eruption on the radiation budget (Minnis et al. 1993), and the subsequent climate response was well predicted (Hansen et al. 1993a; updated in Shine et al. 1994); because of the transient nature of the forcing, this still arguably constitutes the most direct evidence of the linkage between transient forcing and transient response to date. Understanding of ozone RF continued to develop as a result of ongoing analyses of observational data and the advent of (then 2D, latitude–height) chemistry models allowing improved estimates of the longer-term increases in tropospheric ozone (e.g., Wang et al. 1993; Hauglustaine et al. 1994). More sophisticated RF calculations due to sulfate aerosol–radiation interaction were becoming available (e.g., Kiehl and Briegleb 1993; Hansen et al. 1993a; Taylor and Penner 1994), as were the first climate model simulations of aerosol–cloud interaction (Jones et al. 1994). Early attempts to estimate the direct RF from biomass burning (Penner et al. 1992; Hansen et al. 1993a) were presented. Shine et al. (1994) produced the first of IPCC’s many figures of the preindustrial to present-day global-mean forcing incorporating both an estimate of the uncertainty range and a subjective confidence level. Shine et al. (1994) also extended the discussion of the utility of the radiative forcing concept; the chapter included clear demonstrations of the need to include stratospheric temperature adjustment to compute ozone forcings, as IRF and RF could differ in sign.

Climate models were beginning to be used to test the forcing–response relationships for a wide variety of forcings, including the impact of the spatial distribution of forcing; an unpublished study by Hansen et al. (1993b) [a precursor to Hansen et al. (1997b)] reported experiments with an idealized GCM that indicated that extratropical forcings had almost double the impact on global-mean surface temperature change as the same (in the global-mean sense) tropical forcing; ongoing work (e.g., Taylor and Penner 1994) also clearly demonstrated that, while forcing in one hemisphere was felt mostly in that hemisphere, there was still a large nonlocal response. By the time of the SAR update (Schimel et al. 1996), attention had begun to focus on the (positive) direct forcing due to soot (or black carbon) (Chylek and Wong 1995; Haywood and Shine 1995) (see section 5 for details), which highlighted the dependence of the computed forcing on whether the aerosol population was internally or externally mixed. More studies of aerosol indirect forcing were emerging (e.g., Boucher and Lohmann 1995; Chuang et al. 1993) that continued to indicate a significant negative forcing, as well as discussing indirect effects beyond the impact on cloud effective radius. SAR updated the earlier RF figure most notably by splitting the direct effect into its sulfate, biomass burning, and soot components but refrained from giving a central estimate for the aerosol indirect forcing, “because quantitative understanding of this process is so limited.”
CHAPTER 14 R A M A S W A M Y ET AL.

3) IPCC TAR

After four IPCC reports with a focus on RF in the space of just six years, TAR’s analysis (Ramaswamy et al. 2001) was able to assimilate developments over a much longer period, using a much larger body of literature. This was particularly so for tropospheric ozone and aerosol forcing, as a result of many more chemistry transport and GCM studies. These included early studies investigating mineral dust and nitrate aerosols. The sophistication and range of studies on aerosol indirect forcing had increased, with much more effort to separate out first (droplet radii) and second indirect (liquid water path) effects. The associated uncertainty in the first indirect effect could not be reduced beyond that given in SAR; no estimate was given for the second indirect effect because it was “difficult to define and quantify,” but it was noted that it “could be of similar magnitude compared to the first (indirect) effect.” Ramaswamy et al. (2001) also reassessed the simple formulas used by IPCC to compute greenhouse gas RF, which led to a 15% reduction in CO₂ forcing relative to the FAR formula; this and subsequent reports mostly adopted the expressions presented by Myhre et al. (1998a).

TAR also included in its RF summary figure, for the first time, the effect of contrails and contrail-induced cirrus, partly based on work presented in the IPCC Special Report on Aviation and the Global Atmosphere (Prather et al. 1999) and the effect of land-use change on surface albedo.

TAR continued the important discussion on the utility of the RF concept. A larger number of GCM studies, with a more diverse set of forcing mechanisms, were available, leading to the important conclusion that “radiative forcing continues to be a good estimator of global-mean surface temperature response, but not to a quantitatively rigorous extent as in the case of … radiative convective models.” Most notably, Hansen et al. (1997b) had presented a wide-ranging study with a simplified configuration of their climate model that examined the response to both latitudinally and vertically constrained forcings. They showed that forcings confined to specific altitudes could lead to specific cloud or lapse rate responses, and these resulted in marked variations in climate sensitivity for a given forcing. This weakened the perception that the global-mean climate sensitivity for spatially inhomogeneous forcings could be used to determine quantitative aspects of the spatial responses. This important work also presaged later developments in the definition of RF and appears to be the first explicit usage of the concept of “efficacy” (although it was not given that name) that is discussed in the next subsection.

4) IPCC AR4

AR4’s assessment of RF (Forster et al. 2007) brought together many important advances in both its concept and utility of and the quantification of a number of new RF mechanisms. Notably, it was the first report to formally combine all anthropogenic forcings via a Monte Carlo simulation; it concluded that the net anthropogenic forcing since 1750 was “extremely likely” to be positive (central estimate of 1.6 W m⁻²) and that, in the period since 1950, the impact of natural forcings was considered “exceptionally unlikely” to have been comparable to the anthropogenic forcings. A central estimate of the first indirect aerosol forcing (which was labeled the “cloud albedo effect”) was presented (−0.7 W m⁻² with “low level of understanding”).

RF from changes in surface albedo due to black carbon on snow and stratospheric water vapor from CH₄ oxidation were now included on the summary figure; it was noted that the total stratospheric water vapor forcing, based on available observations, could be higher than the methane-only component. In addition to the now-standard IPCC forcing diagram based on changes in concentrations (see summary in Fig. 14-2), an emissions-based version (Fig. 14-4) was presented—this particularly served to highlight the fact that methane emissions (combining the effect of methane change and the indirect forcings from changes in tropospheric ozone, stratospheric water vapor, and CO₂) led to a forcing equivalent to about half that of CO₂. The combined impact of nitrogen oxide (NOₓ) emissions on tropospheric ozone, methane, and nitrate aerosols was found to be negative.

Forster et al. (2007) detailed significant advances in the understanding of the utility of RF. In particular, a number of GCM studies (e.g., Hansen et al. 2005; Shine et al. 2003; Gregory et al. 2004) had explored RF definitions that went beyond the then-standard RF with stratospheric temperature adjustment; this framework allowed for rapid tropospheric adjustments (i.e., those that occur independent of surface temperature change and on time scales of up to a few months) due to changes in clouds, water vapor, and lapse rate to be incorporated in the definition of forcing. These were shown to have greater utility in that the climate sensitivity showed less dependence on the forcing mechanism. This framework would then lead to the AR5 definition of effective radiative forcing [section 2c(5)].

Forster et al. (2007) instead adopted the framework of efficacy, which had been developed in earlier work discussed above, whereby Eq. (14-2) is modified to

\[ dT = E_i \Delta CO_2 dF, \]
where $d\bar{F}$ still represents the (stratospheric-temperature adjusted) radiative forcing and $E_i$ represents the efficacy of a given climate change mechanism, which is given by the ratio of the climate sensitivity for that mechanism to that for CO$_2$; formally, since the CO$_2$ forcing varies slightly with the magnitude of the CO$_2$ change (e.g., Hansen et al. 2005), a more robust definition should be specific to the size of the CO$_2$ perturbation.

The product $E_i d\bar{F}$ was then labeled “effective radiative forcing,” a definition that would be elaborated on in AR5 [section 2c(5)]. A significant number of climate modeling papers had, by then, computed efficacies, with varying levels of agreement; this allowed Forster et al. (2007) to draw tentative conclusions; for example, the combined efficacy for long-lived greenhouse gas forcing was unity, to within 10%; solar forcing, tropospheric ozone, and scattering aerosol forcings had efficacies of 0.7 to 1.0, 0.6 to 1.1, and 0.7 to 1.1, respectively (all with “medium confidence”). There was no consensus on an efficacy for black carbon.

5) IPCC AR5

In AR5, the ERF concept was introduced to allow rapid adjustment processes in the troposphere but avoiding changes that are associated with climate feedbacks (and, in the conventional framework, mediated by surface temperature change—see section 2a) (Boucher et al. 2013; Myhre et al. 2013). The ERF is defined in Myhre et al. (2013) as “change in the net TOA downward radiative flux after allowing for atmospheric temperatures, water vapour and clouds to adjust, but with surface temperature or a portion of surface conditions unchanged.” Figure 14-5, from AR5, summarizes the progression from instantaneous radiative forcing, through RF and ERF, to climate response. AR5 also retained discussion of RF.

No new forcing mechanisms were included in AR5, but the confidence level was raised, relative to earlier IPCC assessments for stratospheric water vapor, aerosol–radiation interactions, surface albedo due to land use, contrails, contrail-induced cirrus, solar irradiance changes, and volcanic aerosols. The only “very low” confidence level was given to rapid adjustment of aerosol–cloud interactions (earlier denoted as aerosol indirect effects). See the summary in Fig. 14-2.

The motivation for introducing the ERF concept was that efficacies [see Eq. (14-3)] for many climate drivers were different to unity when applying RF. This was particularly so for black carbon (Ban-Weiss et al. 2011; Hansen et al. 2005; Ming et al. 2010) and for aerosol–cloud interactions beyond the cloud albedo effect (Twomey effect) (e.g., Lohmann et al. 2010). There was also a growing understanding that rapid adjustments were important for CO$_2$ (Andrews and Forster 2008; Andrews et al. 2012; Doutriaux-Boucher et al. 2009). Furthermore, a clearer distinction between the fast changes (including instantaneous radiative perturbations and the rapid adjustments) and the slow climate feedback processes in terms of their importance for the climate response was elaborated (Andrews et al. 2010; Bala et al. 2010). Importantly, in single model studies, ERF was shown to provide an efficacy much closer to unity than the traditional RF concept (Hansen et al. 2005; Shine et al. 2003). The stratospheric temperature adjustment, which is included in the definition of RF, is also included in ERF. An additional advantage of ERF compared to RF is that a tropopause definition is avoided in the quantification of the forcing (e.g., Shine et al. 2003).
Two methods have been widely adopted to calculate the ERF. One method (Gregory et al. 2004) regresses TOA net radiative imbalance against surface temperature change in coupled climate model simulations. The extrapolation of that regression line to zero surface temperature change then yields the ERF. The second method computes the TOA net radiative fluxes in fixed sea surface temperature (SST) climate model simulations (Hansen et al. 2005); while it is arguably more consistent to fix both land and surface temperatures (Shine et al. 2003), this is difficult to implement in advanced climate models. Instead, Hansen et al. (2005) suggested adjusting the derived ERF to account for the impact of the land surface temperature change on TOA radiative fluxes.

The primary advantage of adopting ERF is that it reduces the level of approximation inherent in Eq. (14-2) across a wide range of climate forcing mechanisms. Nevertheless, there are several limitations associated with its adoption. To some extent, these are reflected in AR5 where the uncertainties in RF of WMGHGs were quantified as 10%, in agreement with earlier IPCC assessments, whereas AR5 assessed WMGHG ERF to have uncertainties of 20%.

The necessity of climate model simulations to calculate tropospheric adjustments makes ERF distinct from either IRF (see section 2b) or RF in several ways. IRF and RF can be quantified using more sophisticated radiative transfer schemes than are typically available in climate models and, for example, can be more easily applied to a wider range of greenhouse gases. In addition, the ERF technique is limited to forcing mechanisms that are of a sufficient size for the impact on TOA fluxes to emerge from the noise of the climate model’s own internal variability [see section 2e(6)].

Since rapid tropospheric adjustment processes are likely to be climate model dependent, this introduces further uncertainties beyond those involved in more traditional forcing definitions. For example, IRF are pure radiative transfer calculations that can be constrained reasonably well with detailed models and a high degree of physical understanding. The stratospheric temperature adjustment that is incorporated in the RF has a well-understood theoretical basis (resulting from the balance between changes in absorption by and emission from the stratosphere). By contrast, tropospheric adjustments are much more complicated. There is less theoretical underpinning with which to constrain these adjustments; this is particularly so for cloud adjustments that result from the complex interplay between different processes that may or may not be well represented in individual climate models. This complicates the distinction between adjustments and feedbacks that are mediated by surface temperature change, and there is no obvious way to quantify the adjustments with observations.

One consequence of these shortcomings is a blurring of the lines between forcings and feedbacks. While the tropospheric adjustments are defined to have a shorter time scale than feedbacks, they also generally involve some coupling to the surface, for example, land warming (in the fixed-SST approach to ERF calculation) or pattern of SST change (in the regression approach). Hence, there is a need to further develop techniques that enable a robust separation of adjustment and feedback processes.

Fig. 14-5. Schematic comparing (a) instantaneous RF, (b) RF, which allows stratospheric temperature to adjust, (c) flux change when the surface temperature is fixed over the whole Earth (a method of calculating ERF), (d) the ERF calculated allowing atmospheric and land temperature to adjust while ocean conditions are fixed, and (e) equilibrium response. [Figure taken from Myhre et al. (2013).]
A specific difficulty is that it is increasingly hard to compare different types of forcing. IRF, which involves purely radiative transfer calculations, has generally not been computed in climate model simulations [see section 2c(6) for future efforts] due to computational considerations; instead, ERF has become the preferred approach to quantifying forcing. Attempts to isolate IRF from ERF using radiative kernels have noted that most of the intermodel spread in ERF from CO₂ forcing does not arise from differences in tropospheric adjustments but rather from differences in IRF (Chung and Soden 2015). Indeed, intermodel differences in the calculation of IRF have been a persistent problem in GCMs (Cess et al. 1993; Soden et al. 2018) despite the presence of accurate and observationally verified line-by-line calculations to constrain their counterparts in climate models (Collins et al. 2006).

6) DEVELOPMENTS SINCE AR5

Marvel et al. (2016) and Shindell (2014) indicate that even for the ERF concept, efficacies [see Eq. (14-3)] may be different for short-lived and regionally heterogeneous compounds like aerosols, ozone, and land use compared to greenhouse gases using various CMIP5 simulations. This would have implications for estimates of climate sensitivity using temperature changes over a historical period. However, the cause of the findings on the efficacies in CMIP5 is under debate (Richardson et al. 2019). The ERF has been further described (Sherwood et al. 2015), and methods to calculate ERF have been better compared (Forster et al. 2016). Forster et al. (2016) find that uncertainties in ERF from fixed SST simulations are much lower than using the regression technique. ERF from the fixed SST simulations can be quantified to an accuracy of 0.1 W m$^{-2}$ at the 5%–95% confidence interval in 30-yr simulations. This implies that ERF from very small forcings (<0.1 W m$^{-2}$) would require many ensembles or very long simulations.

The ERF framework has allowed a clearer understanding of the forcing and climate response of a climate driver; it is now constructive to distinguish into “instantaneous,” “rapid
adjustment,” and “equilibrium.” Figure 14-6 shows how radiative fluxes and surface sensible and latent heat fluxes change for a doubling of CO₂ in multimodel simulations and is a modern version of Fig. 14-3, explicitly using the ERF concept. The instantaneous radiative forcing due to CO₂ is well known to be primarily due to the longwave (LW), with a weak solar [shortwave (SW)] effect (Fig. 14-6, left) shown as the positive values at TOA. Positive values at the surface in Fig. 14-6 indicate that the surface gains energy. If the difference between the flux changes at TOA and the surface is positive, it means the atmosphere gains energy. The SW absorption by CO₂ reduces the solar absorption at the surface and causes a weak change at the TOA, where it is very small compared to LW TOA.

By contrast, the SW contribution to atmospheric absorption is 35%–40% of the LW trapping of energy in the atmosphere. The instantaneous part involves only the initial radiative perturbation. The rapid adjustments (Fig. 14-6, middle) are processes that occur to equilibrate the atmosphere with no SST changes. A doubling of CO₂ gives an initial heating of the troposphere and a cooling of the stratosphere. The cooling of the stratosphere is well known (see section 2b), and the consequent adjustment in the radiative fluxes was included in the early applications of the RF concept. The initial radiative perturbation in the troposphere increases temperature and water vapor and changes clouds. The increase in tropospheric temperature reduces the net atmospheric absorption (giving a radiative cooling), but the reduction in stratospheric temperature has a larger impact on the net atmospheric absorption. The overall rapid adjustments of temperature, water vapor, and clouds lead to an enhanced atmospheric absorption of similar size to the initial heating (Myhre et al. 2018). The atmospheric equilibrium is achieved by reduction in the surface latent and sensible heat fluxes, thereby making a clear link between the atmospheric absorption and precipitation changes, at the global-mean level (e.g., Andrews et al. 2010). The fast response of global-mean precipitation can be estimated, to reasonable accuracy, from the atmospheric component of the ERF (e.g., Samset et al. 2016).

In the full climate response to doubling of CO₂ (Fig. 14-6, right), the surface temperature changes to bring TOA and the atmosphere net fluxes into equilibrium (see section 2a). Note that the initial atmospheric radiative heating (which leads to a precipitation decrease) when ERF is diagnosed turns into a radiative cooling when the full surface temperature response is allowed. Since AR5, there has been an improved quantification of the rapid adjustment processes and their intermodel diversity. Double radiation calls (Ghan 2013) and radiative kernels (Soden et al. 2008) allow a differentiation of the instantaneous radiative perturbation and the rapid adjustment and individual rapid adjustment terms, respectively.

Smith et al. (2018) quantify the rapid adjustment contributions to ERF based on radiative kernels in multimodel simulations for various climate drivers. Figure 14-7 shows the rapid adjustment terms for two scenarios: a doubling of CO₂ and a tenfold increase in the black carbon (BC) abundance. In fact, the IRF at TOA for a doubling of CO₂ and tenfold increase in BC is very similar (Smith et al. 2018), but their total rapid adjustment (the bars on the unshaded background in Fig. 14-7) is strong and of opposite signs. Temperature increases enhance the outgoing longwave radiation and are thus a negative rapid adjustment, which can be seen both for land surface and tropospheric temperature for increase in CO₂ and BC.

The stratospheric cooling due to the CO₂ increase, on the other hand, gives a positive rapid adjustment. Increases in the tropospheric temperature increase water vapor, thus a positive rapid adjustment by increasing the greenhouse effect. For a doubling in CO₂, the high cloud cover increases with a reduction in the lower clouds, but these features are opposite for BC, explaining the different sign of rapid adjustment of clouds for these two climate drivers. For CO₂, the rapid adjustments other than the stratospheric temperature adjustment happen to cancel each other out, making ERF and RF quite similar in that case. Based on available estimates for doubling of CO₂, this was already noted in AR5.

The total rapid adjustment for BC is strongly negative; nevertheless, individual rapid adjustments vary in sign so that the net effect is a residual of these competing effects. Soden et al. (2018) indicate that the large diversity in ERF due to change in CO₂ among GCMs arises from differences in the instantaneous forcing. Results from Smith et al. (2018) support this by having a range of less than 10% (5%–95% confidence interval) of the nonstratospheric temperature rapid adjustments. Combining the uncertainty in the tropospheric rapid adjustment with 10% uncertainty in RF derived from detailed offline radiation schemes as given in AR5 provides an uncertainty in ERF of 14%. Section 10 describes the implication of a reduced uncertainty range of WMGHG forcing compared to that given in IPCC AR5 (20%) for the uncertainty in the total anthropogenic ERF.

**d. Summary and challenges**

A future challenge with respect to the forcing concept is to quantify whether the efficacy is unity when adopting the current definition of ERF for all drivers of climate change and various models and thus to understand the diversity among some previous results (Shine et al. 2013)).
Further, there is a need to better understand the rapid adjustment processes in climate models, both the degree of influence of diversity in IRF as indicated in Smith et al. (2018), but also dedicated process studies comparing GCMs with high-resolution models with weaker degrees of parameterization, such as convection-permitting models. There is a high potential for progress to be made using results from the ongoing CMIP6 model intercomparison project, which is supporting IPCC AR6 (Eyring et al. 2016); efforts have been made to ensure that more diagnostics are available to enable the drivers of ERF to be better quantified and hence for intermodel diversity to be better characterized. Such studies will aid the understanding of whether uncertainties in ERF of CO₂ and other greenhouse gases are substantially larger than using RF, as was indicated in Myhre et al. (2013).

Last, there is a need to develop methodologies to compare weak radiative perturbations, which will continue to need to be quantified using RF, with the major climate drivers that are increasingly being quantified from various model simulations using the ERF concept. It is possible that once the generic understanding of the rapid adjustments has improved, it can be applied to the weak forcings and enable ERF to be estimated from their RFs.

3. CO₂ and other well-mixed greenhouse gases

Accurate computation of the radiative forcing by CO₂ and other gaseous constituents of Earth’s atmosphere only became possible with the advent of accurate spectroscopic measurements in the laboratory. While these measurements still serve as the primary foundation for calculations of the greenhouse effect, for some simple molecules, the combination of quantum and statistical mechanics provides a complementary framework for interpreting the observations and extending them to conditions that have not been directly observed. The inputs to this framework are the energies, numbers (degeneracies), and occupation numbers of excited states of each constituent; the transitions among these excited states; and the interactions of these transitions with light and heat following Einstein’s quantum theory of radiation. One of the first attempts to compute the
effects of doubling CO₂ in an atmosphere in radiative–convective equilibrium produced a remarkably good estimate of 4 K (Hulburt 1931). However, this finding and other supporting evidence was largely unappreciated due to prevailing assumptions prevalent that the strong broadband absorption of water vapor would dominate the regions of the spectrum where CO₂ is radiatively active and that the logarithmic curve of growth of CO₂ radiative effects in these regions would further constrain the impact of rising CO₂ on the climate system.

Parameterizations of the rudimentary laboratory measurements of CO₂ with ambient and prescribed amounts of water vapor accumulated by the mid-twentieth century contradicted the assumption that water vapor would saturate the primary CO₂ bands (Callendar 1941) but did little to change the balance of opinions held by the scientific community. Further, more detailed information on the bands of CO₂ based on its structure and more accurate spectrometers revealed the potential for additional absorption of terrestrial radiation at the edges of these bands and in the upper atmosphere, where the overlap among neighboring absorption lines is greatly reduced (Martin and Barker 1932).

Further progress had to await the need for operational weather forecasting and active and passive remote sensing during and after the Second World War together with the subsequent military investments in computing infrastructure, spectroscopic characterization of Earth’s atmosphere, and the theory of radiative transfer [see historical description by Weart (1997)]. One of the first applications of the new digital computers to atmospheric science revealed that the absorption by CO₂ of upwelling terrestrial radiation in the stratosphere had been systematically underestimated in prior studies based on manual calculations (Kaplan 1952). Subsequent computational solutions to the infrared radiative transfer for the whole atmosphere (Plass 1956c) revealed that the forcing by doubling CO₂ is sufficient to change the mean surface temperature of Earth by about 3.6 K (Plass 1956a).

In the 1960s, general circulation models of the atmosphere were developed that included reasonably complete parameterizations of solar and infrared radiative transfer together with detailed treatments of the radiative properties of Earth’s atmosphere and surface. Pioneering calculations with one-dimensional (Manabe and Wetherald 1967) and three-dimensional (Manabe and Wetherald 1975) models demonstrated the effects of CO₂ on the full climate system and suggested that the surface temperature would increase by approximately 2.5 to 3 K for a doubling of carbon dioxide concentrations (Fig. 14-8). These models and other, simpler energy balance models (Ramanathan et al. 1979) formed the basis for one of the first comprehensive reviews of the state of climate change science for policy makers, the pivotal Charney report (NRC 1979).

Due to rapid increases in computing power, much more exact treatments of radiative transfer known as line-by-line (LBL) codes were developed starting in the 1970s. Early examples include the Fast Atmospheric Signature Code (Drayson 1966; Smith et al. 1978; Fels et al. 1991), a rapid method for computing the Voigt line shape profile that includes line broadening and Doppler shifting (Drayson 1976), and the Automated Atmospheric Absorption Atlas (Scott and Chedin 1981). These codes are based upon comprehensive tabulations of all known absorption lines for over 50 radiatively active gaseous species in Earth’s atmosphere. An early example that has grown significantly and is in widespread usage is the HITRAN database, which was first developed by the Air Force Cambridge Research Laboratories (McClatchey et al. 1973) and subsequently maintained and updated by the Harvard Smithsonian Center for Astrophysics; the most recent version is
HITRAN2016 (Gordon et al. 2017). A parallel European-led line database called Gestion et Etude des Informations Spectroscopiques Atmosphériques (GEISA) was launched in 1974 at the Laboratoire de Météorologie Dynamique (LMD) in France (Chédin et al. 1982; Husson et al. 1992) and was updated most recently in 2015 (Jacquinet-Husson et al. 2016). These LBL codes now serve as the reference radiative transfer codes for calculation of CO2 forcing and its representation in Earth system models (e.g., Collins et al. 2006).

While most climate models compute radiative transfer using full codes, in many applications, simple formulas, based on more detailed calculations, are useful for computing the global annual-mean greenhouse gas radiative forcing from changes in their concentration; such formulas appear in the First Assessment Report of the IPCC (IPCC; Shine et al. 1990) and have been utilized and updated in subsequent IPCC assessment reports. For CO2, to leading order, the forcing in the infrared is proportional to the logarithm of its concentration (Goody and Yung 1995; Pierrehumbert 2011), and in the Third Assessment Report (Ramaswamy et al. 2001), the IPCC formula was augmented to include absorption of solar radiation by CO2 (Myhre et al. 1998a).

One of the lingering uncertainties in the radiative forcing by CO2 is due to remaining uncertainties in its spectroscopic characterization. If we assume that the energies associated with the most important excited states of the CO2 molecule are known precisely, then three principal spectroscopic properties of CO2 involved in computing its radiative forcing are its line strengths, line half widths, and line shapes, in addition to line overlaps with other absorbers. Systematic propagation of these uncertainties through to the radiative forcing from doubling CO2 concentrations suggests that the combined effects of residual errors in these properties are less than approximately 0.7%; this suggests that current LBL models are more than sufficient for accurately computing the climate forcing from this WMGHG (Mlynczak et al. 2016).

a. CH4, N2O, CFCs, and halogenated compounds

Since the atmospheric concentrations of CH4 are two orders of magnitude lower than those of CO2, it was historically difficult to detect through chemical sampling. Methane was first detected under ambient conditions in the 1940s using purely spectroscopic techniques (Migeotte 1948). In turn, the atmospheric concentrations of N2O, the third most important anthropogenic greenhouse gas, are three orders of magnitude lower than those for CO2. The common assumption was that these trace species were present at insufficient levels to have an appreciable impact on the climate system.

As a result, the importance of the radiative forcings by long-lived greenhouse gases other than CO2 went largely unappreciated. Wang et al. (1976) and Donner and Ramanathan (1980) were among the first to compute the impact of increasing concentrations of methane and nitrous oxide and show that the effect could be substantial. Wang et al. (1976) computed the effects of doubling the concentrations of CH4 and N2O (as well as NH3 and HNO3) as a simple proof-of-principle test of anthropogenic perturbations to the concentrations of these compounds.

They found that the combined effects of doubling concentrations of CH4 and N2O also approached 1 K once the climate system had reequilibrated to the elevated concentrations and forcing. To advance beyond these simpler tests required modeling the joint interactions between the physical climate system and the radiatively active species together with the associated networks of chemical sources and sinks for these species. More advanced models that include these interactions can treat the nonlinear effects of spectral overlap among the well-mixed greenhouse gases, ozone, and water vapor. Ramanathan (1980) constructed a prototype of this class of coupled chemistry–climate model and showed that the non-CO2 WMGHGs could contribute nearly half the warming projected for 2025 assuming persistence of extant emissions trends. This conclusion was buttressed in subsequent studies starting with Ramanathan et al. (1985a), who concluded that the minor well-mixed greenhouse gases could contribute as much warming as projected increases in CO2 concentrations. Indirect effects of CH4 and N2O were also uncovered in this time frame and are discussed in greater detail in section 4a.

Because CFCs have strong bands in the midinfrared window, a region of the spectrum otherwise largely transparent to terrestrial infrared radiation, increasing concentrations of these gases leads to rapid increases in Earth’s greenhouse effect (Fig. 14-9). Lovelock et al. (1973), after discovering the ubiquity of CFCs in Earth’s atmosphere, suggested that it might serve as a greenhouse gas. The implications of unchecked historical emissions of these gases (prior to the imposition of the Montreal Protocols in 1987) and the consequent increase in the total greenhouse effect for the mean surface temperature were first calculated by Ramanathan (1975). He found that continuing emissions unabated until the year 2000 would ultimately lead to increases in surface temperature approaching 1 K. Unlike CO2, the major absorption bands of the CFCs are far from saturated, and therefore, the forcing increases linearly and rapidly with increasing concentration.

Ramanathan et al. (1985a) investigated a larger set of compounds, including several CFCs, one HCFC, and
some fully fluorinated compounds (Fig. 14-10). Fisher et al. (1990) expanded the number of halocarbons having a greenhouse effect by providing radiative forcing for a large group of CFC replacements, including HCFCs and HFCs. Some of the halocarbons have major absorption bands outside the midinfrared window and thus have strong overlap with water vapor and even some with CH$_4$ and N$_2$O (Ramanathan et al. 1985a). Pinnock et al. (1995) illustrated how the radiative forcing varies over the infrared spectral region for an increase in 1 ppb of an idealized halocarbon absorbing equally at all wavelengths (Fig. 14-11). The figure shows that halocarbons absorbing particularly in the region 800–1000 cm$^{-1}$ are very efficient compared to, for example, compounds like CF$_4$ with a strong absorption band located closer to 1300 cm$^{-1}$.

Due to the weak absorption by the halocarbons and weak overlap by other gases in the midinfrared window region, Dickinson et al. (1978) showed that these compounds warm the lower stratosphere. This is unlike CO$_2$ and most halocarbons therefore have a positive contribution to radiative forcing from the stratospheric adjustment rather than the negative contribution from CO$_2$ (see section 2). The state of knowledge of radiative forcing per unit concentration change for halocarbons is discussed in section 11c on radiative efficiency. While much of the work to date had stressed the climatic effects of the infrared bands of these minor well-mixed gases, the fact that the shortwave bands also contribute nonnegligible forcing was first highlighted by Collins et al. (2006).

Their study showed that all of the atmosphere–ocean global climate models participating in the Fourth Assessment Report of the IPCC omitted the shortwave effects of CH$_4$ and N$_2$O. This omission was starting to be corrected by the time of the Fifth Assessment, and these effects have now been incorporated into the simple bulk formulas for forcing by these greenhouse gases (Collins et al. 2018; Etminan et al. 2016). The simple formula in Etminan et al. (2016) includes in addition to the direct shortwave effect the shortwave contribution due to stratospheric temperature adjustment and updated water vapor overlap with methane, resulting in a 25% enhancement in the radiative forcing of methane.

b. Summary and challenges

While current LBL models are more than sufficient for accurately computing the climate forcing from WMGHG (Mlynczak et al. 2016), unfortunately, this
The accuracy has not been propagated to the radiation codes used in the ensemble of climate models used for climate projections. The first systematic quantifications of the spread in CO₂ radiative forcing were conducted using the generation of models used in the first IPCC assessment (Cess et al. 1993; Ellingson and Fouquart 1991; Fels et al. 1991), followed by evaluations of modeled forcings used in the fourth (Collins et al. 2006) and fifth (Soden et al. 2018) IPCC assessments.

The 1-sigma relative range in the TOA forcings for the latter two studies is 20%, approximately 1.5 decades larger than the LBL uncertainty (Fig. 14-12). This has significant implications for the interpretation of historical climate change simulations. Reduction and ideally elimination of this large range in CO₂ radiative forcing remains an ongoing challenge for the climate modeling community, with efforts continuing under the WCRP/CMIP6 Radiative Forcing Model Intercomparison Project (RFMIP) (Pincus et al. 2016). Better agreement of ESM radiative parameterizations with LBL models is both feasible and highly desirable. It would help ensure
more accurate interpretations of the historical climate record and more actionable projections of future climate and climate change mitigation scenarios.

4. Short-lived trace gases and chemistry–climate interactions

The importance of atmospheric chemistry for climate began to be recognized after a multitude of advances in the early 1970s revealed a chemically active atmosphere that could be perturbed both by natural and anthropogenic activities (NRC 1984). The foundation for these advances was laid over the preceding four to five decades. This included the knowledge of stratospheric ozone photochemistry (Chapman 1930; Hampson 1965; Bates and Nicolet 1950), urban photochemical air pollution (Haagen-Smit 1952; Altshuller and Bufalini 1965), and tropospheric composition and photochemistry (Cadle and Allen 1970; Bates and Witherspoon 1952). Developments were also dependent on quantitative but (initially) highly uncertain estimates of atmospheric amounts of several trace gases including ozone (in the stratosphere only), nitrogen, oxygen, noble gases, CO₂, H₂O, CH₄, and N₂O (below the tropopause) (see section 1 for more details). We review below developments in atmospheric chemistry that set the stage for the recognition of chemistry as an integral part of the climate system and process-level advances made thereafter that have shaped our knowledge of radiative forcing from short-lived trace gases and chemistry–climate interactions.

a. Atmospheric chemistry and climate connections in 1970s–80s

The importance of chemistry for the radiative forcing of climate change was recognized by studies in the early 1960s that highlighted the important role of ozone in maintaining stratospheric temperature and the tropopause (Manabe and Möller 1961; Manabe and Strickler 1964). The seminal work of Manabe and Wetherald (1967) demonstrated that stratospheric ozone is not only important for maintaining stratospheric temperature but also influences tropospheric and surface temperature, though the effect is small compared to that of CO₂.

A series of new developments in atmospheric chemistry through the 1970s and 1980s established the close links between stratospheric and tropospheric composition and chemistry and how human activities can perturb these linkages with consequences for climate change. The first of these developments was the discovery of human influence on stratospheric ozone through catalytic ozone destruction via NOₓ (Crutzen 1970). Studies showed that increases in NOₓ due to human activities leading to enhanced fossil fuel combustion (such as from a planned fleet of high-altitude supersonic transport planes) (Johnston 1971; Crutzen 1972a; McElroy et al. 1974) or changes in its primary source gas, N₂O (McElroy et al. 1977; Wang and Sze 1980), could cause significant stratospheric ozone loss in the future with consequences for climate (Ramanathan et al. 1976).

Around the same time, Molina and Rowland (1974) identified the role of CFCs as a major source of chlorine responsible for catalytic ozone destruction. Further developments in the understanding of the evolution of stratospheric ozone depletion during this time period are reviewed elsewhere (Wallington et al. 2019; Solomon 1999; Crutzen and Lelieveld 2001).

The next key development in atmospheric chemistry was the identification of the hydroxyl (OH) radical as the primary driver of tropospheric chemistry by Levy (1971). Observational evidence of large concentrations of tropospheric OH (Wang and Davis 1974; Wang et al. 1975; Perner et al. 1976) combined with theoretical and modeling work established that OH plays an extremely important role in controlling the abundance and lifetime of several trace gases emitted at Earth’s surface that were either directly radiatively active (e.g., methane, halogenated hydrocarbons) or affected the abundance of other radiatively active gases by influencing OH concentrations (e.g., CO, NMHCs; McConnell et al. 1971; Wofsy et al. 1972; Singh 1977; Chameides and Cicerone 1978; Logan et al. 1981). Thus, OH came to be recognized as the chemical filter or cleansing agent in the troposphere with important consequences for calculations of radiative forcing.

Finally, a major development that solidified the focus on chemistry–climate interactions was the recognition of the essential role of tropospheric ozone in determining the chemical composition and the radiation budget of Earth’s atmosphere. Till the work of Levy (1971), ozone was assumed to be, except over polluted regions, injected into the troposphere from the stratosphere due to mixing processes, chemically inert, and destroyed at Earth’s surface (Regener 1938; Junge 1962; Fabian and Pruchniewicz 1977). The understanding evolved as both theoretical and observational analysis showed that the photochemical source of tropospheric ozone from OH-initiated oxidation of CO, methane, and other hydrocarbons in the presence of NOₓ dominates over that provided via transport from the stratosphere (Crutzen 1972b, 1973; Chameides and Walker 1973; Fishman et al. 1979b; Fishman and Crutzen 1978). Subsequently, Fishman et al. (1979a) quantified the climate influence of tropospheric ozone using its observed distribution.

Advances in understanding of atmospheric chemistry in both the troposphere and stratosphere, therefore, led to better understanding of its interactions with climate.
Before elaborating on the calculation of ozone radiative forcing, we summarize below the key chemical interactions leading to indirect radiative forcing of climate change that were recognized by the end of the 1980s based on modeling studies with simplistic representation of physical, chemical, and dynamical processes [see Wang et al. (1986), Ramanathan et al. (1987), and Wuebbles et al. (1983) for more details]. Tropospheric hydroxyl radical, as the primary sink for many trace gases, was at the center of these chemistry–climate interactions as illustrated in Fig. 14-13:

- Since reaction with OH is the primary removal mechanism for methane, CO, nonmethane hydrocarbons, many halogenated hydrocarbons, DMS, and SO₂, any changes in the abundance of these gases would alter OH with subsequent feedbacks on the lifetime and abundance and, therefore, climate effects of methane and halogenated hydrocarbons.

- Changes in the emissions of NOₓ, methane, CO, and nonmethane hydrocarbons would affect tropospheric ozone with subsequent climate effects via tropospheric ozone changes or OH-induced changes in the abundances as mentioned above.

- Changes in stratospheric ozone would impact tropospheric OH concentrations by influencing the rate of photolysis of tropospheric ozone resulting in the formation of O(³D)—the primary source of tropospheric OH.

- Increased water vapor in a warmer climate would enhance OH with impacts on abundances of tropospheric ozone (depending on the levels of tropospheric NOₓ) and methane.

- Oxidation of methane is a major source of stratospheric water vapor. Hence, any changes in methane would also influence stratospheric water vapor with subsequent climate implications due to water vapor radiative effects.

By the mid-1980s, the scientific community recognized that a full understanding of the possible changes in ozone distribution and its subsequent effects on climate and biologically important UV radiation would require not only consideration of stratospheric processes but also knowledge of the coupled and nonlinear physical, chemical, and biological processes controlling the chemical composition of the troposphere. This was reflected in the first international scientific assessment of ozone sponsored by the World Meteorological Organization (Ramanathan et al. 1985b).

The inhomogeneous distribution of ozone coupled with the different interactions of stratospheric and tropospheric ozone with solar and longwave radiation necessitated calculations of radiative forcing due to changes in stratospheric and tropospheric ozone separately. The net effect of a reduction in stratospheric ozone on surface temperature depends upon the balance between warming due to enhanced solar radiation reaching the surface–troposphere system and cooling due to reduced longwave radiation emitted by the stratosphere both because of the reduction in stratospheric
ozone and the consequent cooling (the “stratospheric temperature adjustment” process described in section 2). Studies relying on model predictions of ozone distribution showed that the sign of surface temperature change depends on the vertical distribution of the stratospheric ozone change (Ramanathan and Dickinson 1979; Wang et al. 1980; Ramanathan et al. 1985b). Further progress on the sensitivity of surface temperature to observed changes in the vertical distribution of ozone came in 1990 with the iconic work of Lacis et al. (1990), who showed that ozone changes in the upper troposphere and lower stratosphere are most effective in forcing climate change. Surface temperature is much more sensitive to tropospheric ozone perturbations relative to stratospheric ozone changes because the longwave opacity of tropospheric ozone is nearly the same as that of stratospheric ozone, and the solar and longwave effects of tropospheric ozone change affect surface temperature in the same direction (Ramanathan et al. 1985b). Studies estimated the net radiative effect of ozone changes from preindustrial times up to the 1980s to cause warming despite a net reduction in the ozone column (driven by CFC-induced stratospheric ozone depletion) because of the greater radiative efficiency of tropospheric ozone (Owens et al. 1985; Ramanathan and Dickinson 1979).

b. Radiative forcing due to short-lived trace gases: 1990 to 2000

By the early 1990s, a better understanding of the effect of ozone on radiative forcing and its strong dependence on the vertical profile of ozone change throughout the troposphere and stratosphere as well as on its total amount had emerged (Schwarzkopf and Ramaswamy 1993; Lacis et al. 1990; Ramaswamy et al. 1991; Shine et al. 1990). Unlike sparse observational constraints on tropospheric ozone trends, better constraints on stratospheric ozone trends (e.g., Stoalski et al. 1991) facilitated quantitative assessment of its radiative forcing, indicating that ozone reductions between 1979 and 1990 caused a negative radiative forcing (Ramaswamy et al. 1991).

Estimates of radiative forcing due to ozone continued to be refined through the mid-1990s (IPCC 1994), though the level of confidence was low around this time (Fig. 14-2 in section 1). Low confidence in stratospheric ozone forcing was largely driven by more than a factor of 2 spread in model-computed values of stratospheric ozone forcing (Shine et al. 1994, 1995a,b; Schimel et al. 1996). Similarly, for tropospheric ozone, studies based on modeled or limited observations of ozone trends agreed that increases in tropospheric ozone since preindustrial times have resulted in a positive forcing (e.g., Hauglustaine et al. 1994; Mareno et al. 1994; Mohnen et al. 1993), but there was large uncertainty as summarized in the IPCC 1994 special report and SAR (Shine et al. 1994; Schimel et al. 1996). The major difficulty in accurately estimating global ozone forcing was limitations in the knowledge of changes in vertical, horizontal, and temporal distributions of ozone (Prather et al. 1994; Stordal et al. 1995; Ko et al. 1995). This was particularly true for tropospheric ozone, whose distribution was difficult to capture in the two-dimensional models used widely until the mid-1990s [Prather et al. 1994; see Peters et al. (1995) for a review of tropospheric chemistry models until 1995]. Furthermore, it was difficult to demonstrate confidence in model-derived trends because of the lack of strong observational constraints. Chemistry-climate interactions were recognized to have a significant effect on the total radiative forcing of climate and were deemed important to be accounted for in the assessment of potential future climate change as highlighted in chapter 2 of the FAR (Shine et al. 1990) (Fig. 14-13).

The first quantitative estimate of the indirect radiative effects, in terms of global warming potentials (GWPs; see section 11b for definition), from increases in emissions of methane, CO, and NOx was based on results from a two-dimensional model representing the fundamentals of atmospheric chemistry known at the time (Hough and Derwent 1990). Although there was a fair degree of confidence in the sign of the indirect effects (Isaksen et al. 1991), these early estimates were found to be too uncertain and likely overestimated (Johnson et al. 1992; Isaksen et al. 1992; Lelieveld and Crutzen 1992; Isaksen et al. 1991). Two-dimensional tropospheric chemistry models that had been primarily used until the mid-1990s were of limited scope in adequately characterizing the complex chemical and physical processes and the nonlinear interactions between them (Prather et al. 1994; Olson et al. 1997; Stordal et al. 1995) hampering the accurate quantification of indirect radiative forcing (Shine et al. 1994, 1995b; Schimel et al. 1996). Limited atmospheric measurements on a global scale for many species, including ozone, CO, NOx, and NMHCs, needed to characterize historical trends and provide constraints on models, further restricted the ability to robustly quantify indirect radiative forcing.

Progress was however made in better definition and quantification of indirect forcing from methane increases driven by theoretical (Prather 1994) and multi-model analysis (Prather et al. 1994; Stordal et al. 1995). Forcing due to the chemical feedback of methane increases on its own lifetime via reduced tropospheric OH (OH changes discussed in section 4e) was no longer considered an indirect effect as this effect would be implicitly included in the estimates for the forcing due to historical methane changes (Schimel et al. 1996). The
influence of methane increases on tropospheric ozone and stratospheric water vapor was estimated to add about 25% to the direct methane forcing (Schimel et al. 1996; Prather et al. 1994).

From the late 1990s onward, the development and application of sophisticated global climate models with some representation of atmospheric chemistry [see Zhang (2008) and Young et al. (2018) for a historical overview of atmospheric chemistry in global models] combined with improvements in the knowledge of chemical and physical processes affecting the distributions of short-lived gases, and better atmospheric observations led to significant improvements in forcing estimates as assessed in the TAR (Ramaswamy et al. 2001). Results from several studies applying various approaches and observational evidence of ozone loss for a longer period (e.g., MacKay et al. 1997; Forster 1999; Forster and Shine 1997; Hansen et al. 1997a; Granier et al. 1999) enhanced the level of scientific understanding of stratospheric ozone forcing of $-0.15 \pm 0.1 \text{ W m}^{-2}$ for the period 1979 to 1997 (Ramaswamy et al. 2001). Approximate consistency between observed lower-stratospheric temperature trends since the late 1970s and that simulated by global climate models forced with observed ozone losses confirmed that forcing from the decline in ozone is indeed negative (Hansen et al. 1997a). However, a clear attribution was complicated by the possible role of cooling from increasing stratospheric water vapor (Forster and Shine 1999), observations of which were limited.

Results based on global three-dimensional model studies of preindustrial to present-day tropospheric ozone changes driven by precursor emissions (e.g., Roelefs et al. 1997; Haywood et al. 1998; Dorland et al. 1997; Berntsen et al. 1997; Mickley et al. 1999; Brasseur et al. 1998; Stevenson et al. 1998) along with those based on satellite-inferred ozone column changes (Portmann et al. 1997; Kiehl et al. 1999) alleviated uncertainties in the tropospheric ozone forcing estimates (Granier et al. 1999; Ramaswamy et al. 2001). Process studies provided better understanding of the sensitivity of surface temperature to the vertical distribution of ozone and clouds and the spatial distribution of ozone forcing (e.g., Forster and Shine 1997; Hansen et al. 1997a; Hauglustaine and Brasseur 2001).

Although the level of confidence in tropospheric ozone forcing had increased, uncertainties remained because of the large model diversity in predicted historical ozone changes and limited observational constraints on ozone trends (Prather et al. 2001, and references therein). Limitations in our understanding of not only the complex nonlinear chemical interactions between ozone precursors but also the historical evolution of the emissions of specific precursors impeded the quantitative attribution of ozone forcing up until this time (Prather et al. 2001, and references therein).

Studies also highlighted the importance of including feedbacks between climate and chemistry on the assessment of the climate impact of short-lived species (Prather et al. 2001). Here, we do not cover the details of this feedback but refer to past IPCC reports and several review papers on this topic (Prather et al. 2001; Jacob and Winner 2009; Fiore et al. 2012, 2015; Isaksen et al. 2009; Denman et al. 2007; Von Schneide messenger et al. 2015; Brasseur 2009; Kirtman et al. 2013; Monks et al. 2015).

Much progress was made in the quantitative estimates of forcings from chemistry–climate interactions in the latter half of the 1990s as assessed in the TAR. The indirect forcing from methane changes continued to be the best studied, with explicit quantification of the individual effects on its own lifetime, tropospheric ozone, stratospheric water vapor, and CO$_2$ (e.g., Hauglustaine et al. 1995; Lelieveld et al. 1998; Fuglestvedt et al. 1996). Modeling studies also quantified indirect forcing from changes in CO and a suite of NMHCs through their influence on methane lifetime, tropospheric ozone, and CO$_2$ (Daniel and Solomon 1998; Johnson and Derwent 1996).

Accurate calculation of the indirect forcing of NO$_x$ remained challenging because of counteracting effects—increased NO$_x$ emissions not only increase tropospheric ozone, but also increase OH concentration. The former produces a short-lived regional positive forcing while the latter lowers methane abundance (with a consequent decrease in ozone), which produces a longer-lived global negative forcing. This longer-lived global forcing partially offsets the short-lived regional ozone forcing (Ramaswamy et al. 2001, and references therein). Studies showed that the ozone and OH perturbations strongly depended on the location of NO$_x$ emission perturbations because of the nonlinear ozone chemistry and differences in mixing regimes (e.g., Fuglestvedt et al. 1999).

Diversity in results of model studies that resolved the complex and nonlinear effects of emission changes on ozone and OH radical and limitations in observational constraints to build confidence in them remained a significant source of uncertainty in these estimates of indirect forcings (Ramaswamy et al. 2001).


Over the past two decades, the development of increasingly sophisticated comprehensive global chemistry models in terms of their design (e.g., models with coupled stratospheric–tropospheric chemistry, global climate model with online chemistry) and the representation of complex physical and chemical processes (e.g., trace gas-aerosol
interactions, interactive natural emissions), combined with better estimates of trace gas emissions and the availability of longer observational records has facilitated advances in the attribution of the changes in short-lived trace gases and their forcings (Forster et al. 2007; Myhre et al. 2013).

Consideration of coupled stratospheric and tropospheric chemistry in global models has facilitated greater understanding of the influence of changes in stratospheric ozone and ozone-depleting substances (ODSs) on tropospheric ozone and the effect of tropospheric ozone precursors on stratospheric ozone (e.g., Shindell et al. 2006; Hegglin and Shepherd 2009; Eyring et al. 2013; Young et al. 2013), which has led to better accounting of these impacts on the radiative forcing due to ozone (Gauss et al. 2006; Stevenson et al. 2013; Myhre et al. 2013; Forster et al. 2007; Shindell et al. 2013a; Banerjee et al. 2016). Coupled chemistry–climate models have enabled assessment of the changes in ozone induced by climate change (i.e., chemistry–climate feedbacks; see Isaksen et al. 2009; more references), and the resulting radiative forcing (e.g., Gauss et al. 2006; Stevenson et al. 2013; Forster et al. 2007) and feedbacks on climate (e.g., Nowack et al. 2015; Chiodo et al. 2018; Marsh et al. 2016).

A major development in the quantification of forcing due to tropospheric ozone and chemistry–climate interactions over this period has been the adoption of an emissions-based approach (Shindell et al. 2009, 2005) to estimate the contribution of anthropogenic emissions of individual ozone (or aerosol) precursors to the pre-industrial to present-day radiative forcing either via direct influences (e.g., ozone, methane, or aerosols) or indirect effects (Myhre et al. 2013; Forster et al. 2007). With this approach, model representation of couplings between gas-phase and aerosol chemistry in the troposphere has helped elucidate indirect effects of trace gases on aerosols via influences on ozone and OH and the resulting forcing (Shindell et al. 2009; Von Schneidemesser et al. 2015). Additionally, indirect forcing through detrimental ozone effects on vegetation (see section 6 for discussions on forcing from land and biogeochemical interactions) has also been explored (Sitch et al. 2007; Collins et al. 2010; Kvalevåg and Myhre 2013). These emissions-based radiative forcing estimates (Fig. 14-14) give a significantly different relative importance to various emissions (Forster et al. 2007; Myhre et al. 2013) than that suggested by abundance-based assessments in the past. Radiatively active short-lived trace gases (and aerosols; see section 5) and their precursors are now collectively termed as short-lived climate forcers (SLCFs) as their climate impact is mainly felt within the first one to three decades (near term) of their emissions (Myhre et al. 2013; Fiore et al. 2015) in contrast to long-lived greenhouse gases. Furthermore, the short lifetimes of SLCFs result in spatially inhomogeneous abundances and associated forcings highly sensitive to the location of emissions. Consequently, climate influence from SLCFs is more important on a regional scale (e.g., Fry et al. 2012; Collins et al. 2013; Aamaas et al. 2017), contrary to the relatively homogeneous spatial influence from well-mixed greenhouse gases.

The question of how global-mean hydroxyl radical has evolved in the past and will change in the future in response to anthropogenic emission and climate change remains highly relevant to the estimates of SLCF radiative forcing given the dependence of SLCF atmospheric lifetimes on OH (section 4a). Significant progress has been made in the understanding of fundamental atmospheric chemistry of OH with advances in both observations and modeling (e.g., Stone et al. 2012; Rohrer et al. 2014); however, the answer to this question remains at an impasse. The atmospheric chemistry community has mostly relied on global chemistry models to derive past changes and predict future evolution of OH over long time scales, and on proxies, such as methyl chloroform, to derive OH variability over the past ∼35 years during which we have observations (e.g., Prinn et al. 2001; Bousquet et al. 2005; Montzka et al. 2011; Rigby et al. 2017; Turner et al. 2017). There is no consensus in the global model estimates of changes in tropospheric-mean OH abundance from preindustrial times to the present day based on studies over the past ∼40 years as displayed in Table 14-1. The simulated change in present-day OH relative to preindustrial times ranges from a decline to no change to an increase due to varying levels of offsetting effects from increases in OH sinks (methane, CO, NMHCs) and increases in factors that increase OH (water vapor, tropospheric ozone, NOx, and UV radiation) (e.g., Naik et al. 2013). This is in contrast to the 30% decline in present-day OH relative to preindustrial inferred from ice core measurements, although with large uncertainties (Alexander and Mickley 2015). There are also large discrepancies in the projections of global OH levels with implications for estimates of future SLCF forcing (e.g., Voulgarakis et al. 2013). Changes in OH over the past ∼35 years and their role in the renewed growth in atmospheric methane since 2007 are intensely debated in the literature, with no consensus view (Turner et al. 2019).

Global chemistry–climate models have remained the tools of choice to quantify the contribution of SLCF emissions to radiative forcing of climate change as observational constraints are sparse (e.g., for tropospheric ozone; Bowman et al. 2013). Multimodel intercomparison projects (MIPs) involving coordinated experiments with chemical models provide a means of exploring structural uncertainty related to model representation.
of various physical and chemical processes determining the distribution and budgets of SLCFs and have informed IPCC as well as other international assessments [see Young et al. (2018) for a brief history of MIPs for chemistry].

Similar to climate model intercomparisons (e.g., Meehl et al. 2007), analysis is focused on multimodel means because the ensemble average across structurally different models shows better agreement with available observations with individual model biases canceling out, while the spread across models is considered a measure of uncertainty (e.g., Young et al. 2013). However, because these ensembles represent “ensembles of opportunity,” the spread across models does not necessarily span the full range of structural as well as process uncertainty (Tebaldi and Knutti 2007).

For chemistry models, early MIPs focused on exploring the uncertainty in model representation of

![Figure 14-14](image-url)

**Figure 14-14.** Best estimate of global average radiative forcing for the period 1750 to 2011 for emitted chemical species (well-mixed and short-lived gases, aerosols, and their precursors) or other factors. As shown in the inset in the top portion of the figure, many colors are used to represent RF from emitted species affecting several chemicals while red (positive forcing) and blue (negative forcing) are used for emitted components that affect fewer forcing agents. The vertical bars represent the relative uncertainty of the RF induced by each component. [From Myhre et al. (2013).]
specific processes affecting the distribution and budget of tropospheric ozone and related trace gases [e.g., Photochemical Model Intercomparison (PhotoComp) in Olson et al. (1997) and Tropospheric Oxidant Model Comparison (OxComp) in Prather et al. (2001)]. Computation of ozone radiative forcing within MIPs came about later in the 2000s, beginning with the framework of Atmospheric Chemistry Composition Change: The European Network (ACCENT; Gauss et al. 2006) that informed the AR4 report (Forster et al. 2007). The specifications of the simulations for MIPs improved with the development of a consistent set of gridded anthropogenic precursor emissions describing their preindustrial to present-day evolution (Lamarque et al. 2010). This common dataset employed by the more recent Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIIP) (Lamarque et al. 2013) allowed for increased comparability of model simulations of tropospheric ozone (and aerosols) abundances and resulting radiative forcings as assessed in AR5 (Myhre et al. 2013). Uncertainties in emission estimates (e.g., Granier et al. 2011; Bond et al. 2013) have consequences for SLCF radiative forcing. New and revised estimates of the historical evolution of SLCF and their precursor emissions (Hoesly et al. 2018; van Marle et al. 2017) provide a means of exploring the contribution of emission uncertainty to SLCF forcing uncertainty. The Aerosol Chemistry Model Intercomparison Project (AerChemMIP) in support of the forthcoming IPCC assessment (AR6) is designed to quantify and explore uncertainties in the forcing due to anthropogenic emissions of SLCFs, thereby providing better constraints on the role of SLCFs in climate forcing (Collins et al. 2017).

d. Summary and challenges

In this section, we have reviewed the evolution of our knowledge of radiative forcing from short-lived trace gases and chemistry–climate interactions over approximately the past four decades.

Significant progress has been made beginning with the recognition of the role of stratospheric ozone on climate change to the scientific understanding and quantitative estimate of the contribution of emissions of a suite of SLCFs to Earth’s radiative forcing. The use of comprehensive global chemistry–climate models combined with observational constraints where available have enhanced our ability to capture complex chemical interactions in the computation of SLCF radiative forcing. However, challenges remain in quantifying the forcing

<table>
<thead>
<tr>
<th>Reference</th>
<th>% change in OH since preindustrial</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>McElroy (1989)</td>
<td>+60%</td>
<td>1D model</td>
</tr>
<tr>
<td>Hough and Derwent (1990)</td>
<td>−19%</td>
<td>2D model</td>
</tr>
<tr>
<td>Valentin (1990)</td>
<td>−9%</td>
<td>2D model</td>
</tr>
<tr>
<td>Law and Pyle (1991)</td>
<td>−13%</td>
<td>2D model</td>
</tr>
<tr>
<td>Pinto and Khalil (1991), Lu and Khalil (1991)</td>
<td>−4%</td>
<td>1D model, multi-1D model</td>
</tr>
<tr>
<td>Staffelbach et al. (1991)</td>
<td>−30%</td>
<td>Ice core measurements of formaldehyde</td>
</tr>
<tr>
<td>Crutzen and Zimmerman (1991)</td>
<td>−10% to −20%</td>
<td>3D model</td>
</tr>
<tr>
<td>Thompson et al. (1993)</td>
<td>−20%</td>
<td>Multi-1D model</td>
</tr>
<tr>
<td>Martinerie et al. (1995)</td>
<td>+6%</td>
<td>2D model</td>
</tr>
<tr>
<td>Berntsen et al. (1997)</td>
<td>+6.8%</td>
<td>3D model</td>
</tr>
<tr>
<td>Roelofs et al. (1997)</td>
<td>−22%</td>
<td>3D model</td>
</tr>
<tr>
<td>Brasseur et al. (1998)</td>
<td>−17%</td>
<td>3D model</td>
</tr>
<tr>
<td>Wang and Jacob (1998)</td>
<td>−9%</td>
<td>3D model</td>
</tr>
<tr>
<td>Mickley et al. (1999)</td>
<td>−16%</td>
<td>3D model</td>
</tr>
<tr>
<td>Grenfell et al. (2001)</td>
<td>−3.9%</td>
<td>3D model, NMHCs</td>
</tr>
<tr>
<td>Hauglustaine and Brasseur (2001)</td>
<td>−33%</td>
<td>3D model</td>
</tr>
<tr>
<td>Shindell et al. (2001)</td>
<td>−5.9%</td>
<td>3D model</td>
</tr>
<tr>
<td>Lelieveld et al. (2002)</td>
<td>−5%</td>
<td>3D model</td>
</tr>
<tr>
<td>Lamarque et al. (2005)</td>
<td>−8%</td>
<td>3D model</td>
</tr>
<tr>
<td>Shindell et al. (2006)</td>
<td>−16%</td>
<td>3D model</td>
</tr>
<tr>
<td>Sofen et al. (2011)</td>
<td>−10%</td>
<td>3D model</td>
</tr>
<tr>
<td>John et al. (2012)</td>
<td>−6%</td>
<td>3D model</td>
</tr>
<tr>
<td>Naik et al. (2013)</td>
<td>−0.6% ± 8.8%</td>
<td>Multi-3D model</td>
</tr>
<tr>
<td>Murray et al. (2014)</td>
<td>+7.7% ± 4.3%</td>
<td>Multi-3D model</td>
</tr>
<tr>
<td>Achakulwisut et al. (2015)</td>
<td>−8% to +17%</td>
<td>Multi-3D Model</td>
</tr>
</tbody>
</table>
due to anthropogenic emissions of SLCFs, as outlined below:

- A persistent uncertainty in constraining the radiative forcing from SLCFs is the limited or lack of knowledge of preindustrial precursor emissions and atmospheric burdens [e.g., for tropospheric ozone as highlighted by Stevenson et al. (2013)].

- The spatial distribution of ozone precursor emissions has undergone a dramatic change over the last couple of decades, with emissions declining in the developed midlatitude and rising in the developing tropical regions (e.g., Zhang et al. 2016). The consequences of such emission distribution for chemistry–climate interactions and consequent SLCF forcing is not clear.

- The debate over how global-mean OH is changing in response to changing anthropogenic emissions and climate change and the implications of this change for the abundance and lifetime of SLCFs is yet to be resolved. It has been a challenge to narrow down the reasons for differences in global model simulations of the evolution of atmospheric OH (e.g., Naik et al. 2013; Voulgarakis et al. 2013). Recent efforts combining observations and model results in novel ways show promise in understanding the causes of model disagreement (e.g., Nicely et al. 2017; Prather et al. 2018).

5. Tropospheric aerosols

Aerosols scatter and absorb radiation (the direct effect), and also act as cloud condensation nuclei whereby they modify the microphysical and macrophysical properties of clouds (the indirect effect). Increased concentrations of aerosols from anthropogenic activity therefore exert a radiative forcing of climate. The importance of atmospheric aerosols had been long established in the areas of atmospheric visibility (Koschmeider 1924) and human health (e.g., Lippmann and Albert 1969), but aerosols were originally considered to be of only minor consequence in terms of their impact on climate via direct and indirect effects (Twomey 1959; McCormick and Ludwig 1967; Bolin and Charlson 1976). Simple models of the impact of aerosols on planetary albedo in terms of their absorptance and the reflectance of the underlying surface had been developed (e.g., Ensor et al. 1971; Reck 1974; Chylek and Coakley 1974), but the radiative forcing was not quantified owing to the lack of knowledge of the anthropogenic aerosol perturbation.

Observational evidence of aerosol–cloud interactions was hypothesized from observations of ship tracks after satellite sensors were launched (Conover 1966), but quantifying their radiative effect was only possible once detailed spectral information from satellites became available (Coakley et al. 1987). Observations of surface solar insolation suggested a widespread reduction in irradiance at the surface (Stanhill and Moreshet 1992), although the causes were difficult to attribute. Thus, until the late 1980s, aerosols were considered insignificant in terms of radiative forcing when compared to that from changes in atmospheric concentrations of greenhouse gases. This view changed in the early 1990s. Global chemical transport models (CTMs) that were able to model the aerosol life cycle of emission, chemical transformation, transportation, and deposition and hence model the anthropogenic perturbation to aerosol concentrations were combined with relatively simple radiative transfer models.

a. Simple models of aerosol–climate interactions: The early 1990s

Sulfate aerosol was the first aerosol species to be comprehensively investigated owing to the anthropogenic emissions of the SO$_2$ gaseous precursor being $\sim$100 Tg yr$^{-1}$, that is, exceeding natural emissions by around a factor of 5 (Langner and Rodhe 1991). These early global CTMs typically had a spatial resolution of $1^\circ \times 1^\circ$ latitude/longitude with $\sim$10 coarsely spaced atmospheric levels in the troposphere (Zimmermann et al. 1989). The impact of anthropogenic emissions of sulfur dioxide and the resulting sulfate aerosols on the radiative forcing of Earth’s climate was initially quantified by Charlson et al. (1991, 1992), who used a multiple-scattering approximation to derive an equation for the change in the planetary albedo owing to a purely scattering aerosol and focused on the change in cloud-free regions.

This related the direct radiative forcing of sulfate aerosol, $F_{\text{direct}}$, to the total solar irradiance $S_0$, the atmospheric transmission $T_{\text{at}}$, the cloud fraction $A_c$, the surface reflectance $R_s$, the aerosol single-scattering albedo $\omega_a$, the fraction of light backscattered to space $\beta_{\text{aer}}$, and the perturbation to the aerosol optical depth since preindustrial times $d\tau_{\text{aer}}$:

$$F_{\text{direct}} = \frac{1}{2} S_0 T_{\text{at}}^2 (1 - A_c)(1 - R_s)^2 \beta_{\text{aer}} d\tau_{\text{aer}}.$$

Charlson et al. (1992) also derived an expression for the Twomey effect, which is the aerosol impact on the cloud droplet effective radius under the assumption of constant cloud liquid water (Twomey 1977). Charlson et al. (1992) used global-mean estimates of the various parameters coupled to newly available estimates of the perturbation to the total aerosol concentrations caused by anthropogenic emissions (Langner and Rodhe 1991) and concluded that the resulting global-mean radiative forcing for aerosols (preindustrial times to circa 1980s) was $F_{\text{direct}} = -1.3$ W m$^{-2}$ and $F_{\text{twomey}} = -1$ W m$^{-2}$ with significant uncertainty owing to the neglect of...
subsequent impacts on cloud liquid water (i.e., the Albrecht effect; Albrecht 1989), the simplicity of the models used, and the lack of account of spatial correlation between the various parameters. Penner et al. (1992) used a similar method to derive an initial estimate of the radiative forcing due to biomass burning aerosols from combined direct and indirect effects as strong as −2 W m⁻².

Simple representations of the direct radiative forcing of sulfate aerosol were straightforward to implement in fully coupled ocean–atmosphere models (Mitchell et al. 1995) because, for cloud-free regions, the local surface reflectance could simply be increased in proportion to \( dR_s = (1 - R_s) \beta_{aer} dT_{aer} \mu_0 \).

Implementing this parameterization in the Met Office Climate Model showed a reduced rate of warming particularly in the Northern Hemisphere in their climate model simulations, bringing the simulated surface temperature change into better agreement with observations.

Because the radiative forcing due to aerosols could conceivably outweigh that of increased concentrations of well-mixed greenhouse gases, there were significant efforts to better quantify aerosol radiative forcing.

### b. Refinement of aerosol direct and indirect effect modeling studies: The mid-1990s

Simple models continued to play a significant role. Early simple theoretical models of radiative impacts of partially absorbing aerosols (e.g., Ensor et al. 1971; Reck 1974; Chylek and Coakley 1974) were now extended (Haywood and Shine 1995; Chylek and Wong 1995) by accounting for aerosol absorption via the aerosol single-scattering albedo \( \omega_0 \), day-length fraction \( D \), and spatially resolved parameter values rather than global-mean values:

\[
F_{direct} = -DT_{at}^2(1 - A_c)[\omega_0 \beta_{aer} (1 - R_s)^2 - 2(1 - \omega_0)R_s] dT_{aer}.
\]

Assuming a mass fraction between absorbing BC and scattering sulfate based on in situ measurements weakened the sulfate \( F_{direct} \) from −0.34 W m⁻² [already much weaker than that diagnosed by Charlson et al. (1991, 1992)] to between −0.10 and −0.30 W m⁻² depending on the assumed sulfate climatology and mixing state and that there was no direct radiative forcing in cloudy areas. This study established the importance of aerosol absorption, particularly when it was recognized that the positive radiative forcing impact would be amplified if absorbing aerosols resided above underlying cloud (e.g., Haywood and Shine 1997). Regionally, \( dF_{direct} \) can be either positive (for low \( \omega_0 \) and high \( R_s \) or above reflective cloud) or negative (for high \( \omega_0 \) and low \( R_s \)) as demonstrated by the “real color” image (Fig. 14-15).

A more comprehensive estimate of \( F_{direct} \) for sulfate aerosol was performed by Kiehl and Briegleb (1993), who imposed monthly mean climatologies of sulfate mass burden (Langner and Rodhe 1991), an aerosol size distribution, and suitable refractive indices to derive aerosol optical properties. These optical properties were then included in offline radiative transfer calculations using meteorological fields from observations. The \( F_{direct} \) of sulfate was evaluated as being a modest −0.3 W m⁻² [this affirmed the simple model of Haywood and Shine (1995)]. Similarly, previous results for biomass burning aerosol were regarded as too strongly negative owing to lack of account of aerosol absorption and preindustrial biomass burning and the lack of a discernible cooling trend in the climatic record. GCM investigations of the aerosol indirect effect were also performed (Jones et al. 1994; Boucher and Lohmann 1995) using the model-simulated clouds.

Climatological sulfate concentrations (Langner and Rodhe 1991) were utilized together with parameterizations based on airborne observations that related the cloud droplet effective radius to the aerosol number concentration in marine and continental environments (Martin et al. 1994). These simulations indicated an \( F_{Twomey} \) of −1.3 W m⁻² (Jones et al. 1994) and −0.5
to $-1.5 \text{ W m}^{-2}$ (Boucher and Lohmann 1995), but the uncertainty remained significant. Boucher (1995) made a first estimate using satellite observations of the difference between interhemispheric cloud effective radius (Northern Hemisphere: 11.0 $\mu$m; Southern Hemisphere: 11.7 $\mu$m; Han et al. 1994) but acknowledged that the contribution of aerosols from natural land surfaces made the results difficult to interpret in the context of anthropogenic radiative forcing. Recognizing the fidelity of the refinements, IPCC (Schimel et al. 1996) suggested a best estimate for $F_{\text{direct}}$ of $-0.5 \text{ W m}^{-2}$ (range $-0.25$ to $-1.0 \text{ W m}^{-2}$), which was derived from a combination of the radiative forcing of sulfate ($-0.4 \text{ W m}^{-2}$), biomass burning ($-0.2 \text{ W m}^{-2}$), and fossil fuel black carbon (FFBC; $+0.10 \text{ W m}^{-2}$). IPCC (Schimel et al. 1996) recognized that a best estimate of $F_{\text{Twomey}}$ was impossible to establish without further model simulations and observational constraints and suggested a range of 0 to $-1.5 \text{ W m}^{-2}$.

t. The proliferation of GCM-based estimates and the requirement for validation data: Late 1990s to early 2000s

The development of global model-based estimates of aerosol species other than sulfate continued apace. Tegen and Fung (1995) developed a global model of mineral dust and highlighted that, in addition to impacts in the solar region of the electromagnetic spectrum, coarse-mode aerosols can have a significant impact by absorbing and reemitting terrestrial radiation. Any anthropogenic fraction of mineral dust was recognized as being very uncertain. A first black carbon climatology was also produced (Cooke and Wilson 1996), reiterating that global black carbon emissions ($-14 \text{ Tg yr}^{-1}$; cf. SO$_2$ at $-100 \text{ Tg yr}^{-1}$) would lead to anthropogenic aerosol that was partially absorbing, that is, gray rather than white (Fig. 14-15). The first estimates of nitrate aerosol direct radiative forcing were also produced (van Dorland et al. 1997) but were highly uncertain owing to differences in the partitioning between the accumulation and coarse modes (Adams et al. 2001; Jacobson 2001). The recognition that the different aerosol types needed to be represented for accurate determination of total aerosol radiative forcing led to a rapid expansion of GCM estimates based on aerosol climatologies derived from global CTMs.

Aerosol optical properties are determined by the (wavelength dependent) refractive index of the particles and the particle size distribution. Recognizing that aerosol direct effects needed to be represented using more flexible radiative transfer codes that allowed integration over the full solar spectrum and range of solar zenith angles led to a comprehensive multimodel radiative transfer intercomparison for sulfate aerosol (Boucher et al. 1998).

This study showed a relatively modest variation in radiative effect between the radiative transfer models, indicating that the radiative transfer codes of reduced complexity in GCMs, but more refined than those used in the earlier simplified model calculations, could adequately describe aerosol direct radiative effects.

It was recognized that GCMs were a suitable tool for allowing representation of the variability in humidity, surface reflectance, aerosol, and cloud but computational expense meant that CTMs were used for computing, for example, monthly mean distributions of sulfate aerosol, and these monthly mean fields were then input to the GCMs, which computed the direct and indirect effects using their internal radiative transfer models (e.g., Kiehl and Briebble 1993; Boucher and Anderson 1995; Boucher and Lohmann 1995; Kiehl and Rodhe 1995; Haywood et al. 1997; Haywood and Ramaswamy 1998; Hansen et al. 1998). However, CTMs and GCMs were increasingly combined so that the sulfur chemistry, transport, deposition, and direct and indirect radiative forcing could be explicitly calculated. This method had the benefit that aerosol concentrations could be precisely correlated with fields determining aerosol production and removal, that is, clouds and precipitation (e.g., Graf et al. 1997; Feichter et al. 1997; Myhre et al. 1998b; Iversen et al. 2000; Ghan et al. 2001; Jacobson 2001; Jones et al. 2001). Nevertheless, the limitations of model resolution were recognized; GCMs with their coarse resolution of approximately hundreds of kilometers were unable to represent the subgrid-scale details such as relative humidity and detailed distributions of gas-phase and aqueous-phase production of sulfate aerosol (Ghan et al. 2001).

Direct radiative forcing calculations were made for fossil fuel BC (Haywood and Ramaswamy 1998; Penner et al. 1998; Cooke et al. 1999), fossil fuel organic carbon (Penner et al. 1998; Cooke et al. 1999), biomass burning aerosol (Penner et al. 1998; Iacobellis et al. 1999), total BC (Hansen et al. 1998; Haywood and Ramaswamy 1998; Jacobson 2001), and fossil and biomass burning organic carbon (Hansen et al. 1998; Jacobson 2001). These models treated each of the aerosol types separately (i.e., an external mixture), although some of these models began to represent multicomponent aerosols as internal mixtures, which can have particular relevance for inclusion of absorbing BC cores within scattering shells (e.g., Jacobson 2001) producing a “lensing” effect that enhances the absorption (e.g., Lesins et al. 2002).

GCM studies of indirect effects tended to rely on empirical relationships between aerosol number (e.g., Jones et al. 1994) or mass (Boucher and Lohmann 1995).
and cloud droplet number concentrations (CDNC), but prognostic mechanistic parameterizations that attempted to explicitly account for aerosol activation and cloud nucleation began to appear (e.g., Lohmann et al. 2000; Ghan et al. 2001). By contrasting polluted and unpolluted clouds, comprehensive aircraft-based observational measurement campaigns [e.g., Second Aerosol Characterization Experiment (ACE-2); Brenguier et al. 2000] were able to show clear evidence of aerosol Twomey effects, but definitive evidence of Albrecht effects remained elusive.

Until this point, there was little/no information available from observational sources with a global reach (i.e., satellites or global surface networks) with which the global models could be challenged. The first satellite retrievals of \( \tau_{aer} \) (at midvisible wavelength) appeared, based on the reflectance from a single visible spectral channel from the AVHRR satellite sensor (Husar et al. 1997; Fig. 14-16).

These retrievals were restricted to cloud-free regions over ocean owing to difficulties in accurately characterizing surface reflectance properties over land and cloudy regions, but for the first time, these retrievals were able to detect the geographic distribution of aerosols and how these distributions shifted according to the season. These observations emphasized that, to compare model results against those from observations, both natural and anthropogenic aerosols need to be modeled, particularly those of mineral dust (e.g., Woodward 2001) and sea salt aerosol (Fig. 14-16).

Improved detection algorithms soon followed using two channels (AVHRR; Mishchenko et al. 1999) or polarization (POLDER; Deuzé et al. 2001). The use of two-channel retrievals and polarization allowed, for the first time, separation between coarse- and fine-mode aerosol particles based on the measured Angstrom exponent (i.e., the wavelength dependence of \( \tau_{aer} \)) and depolarization, respectively. Initial estimates of the direct radiative effect were also made over the cloud-free oceans (e.g., Haywood et al. 1999; Boucher and Tanré 2000), but the problem of deriving a radiative forcing (i.e., the change in the radiative effect since preindustrial times) remained. Concurrently with satellite observations, significant investment was made in the global network of aerosol sun photometers (Holben et al. 1998) that were able to measure \( \tau_{aer} \) from direct sun measurements. The first sun photometer was deployed in 1993, but this network was to blossom over the next two decades with the number of sites operational in June 1998/2008/2018 increasing from 33/200/400 globally. AERONET has become a mainstay for checking both the calibration of satellite \( \tau_{aer} \) retrievals and modeled \( \tau_{aer} \).

Given the rapid growth in model estimates of direct and indirect effects, IPCC (2001) commissioned an intercomparison workshop (Penner et al. 2001) to provide global model estimates of various aerosol parameters such as speciated natural and anthropogenic burdens and direct and indirect radiative forcings; in many ways, this may be considered the forerunner of the Aerosol Comparisons between Observations and Models project (AeroCom; Kinne et al. 2006).

Because of the rapid expansion of estimates of both direct and indirect effects, IPCC (Ramaswamy et al. 2001) expanded the number of aerosol species assigned a direct radiative forcing: sulfate (\(-0.40 \text{ W m}^{-2}, \times 2\) uncertainty), biomass burning aerosols (\(-0.20 \text{ W m}^{-2}, \times 3\) uncertainty), FFBC (\(+0.2 \text{ W m}^{-2}, \times 2\) uncertainty), fossil fuel organic carbon (FFOC; \(-0.10 \text{ W m}^{-2}, \times 3\) uncertainty), mineral dust (range \(-0.6 \text{ to } +0.40 \text{ W m}^{-2}\)), and estimated the total combined \( F_{\text{direct}} \) to be \(-0.6 \text{ W m}^{-2}\) (range \(-0.1 \text{ to } 1 \text{ W m}^{-2}\)). While diagnosing aerosol direct radiative effects within GCMs was straightforward, diagnosing aerosol indirect forcing beyond the Twomey effect became problematic because the strict definition of radiative forcing required “surface and tropospheric
temperatures and state held fixed at the unperturbed values” (Ramaswamy et al. 2001). This definition precluded allowing changes in cloud macrophysical properties such as cloud liquid water path (Albrecht 1989) and subsequent impacts on cloud fraction, cloud height, etc. Thus, in a strict radiative forcing sense, only \( F_{\text{Twomey}} \) could be calculated within GCMs, although some studies simply used the difference between two simulations with preindustrial and present-day aerosols with fixed sea surface temperatures to attempt to diagnose aerosol–cloud impacts beyond Twomey effects (Rotstayn 1999; Jones et al. 2001). These studies can be thought of as forerunners of the “effective radiative forcing concept” (section 2). The fidelity and utility of ERF was far from proven however, so IPCC (2001) assigned only the Twomey effect with a range of uncertainty. A comprehensive discussion of the basis and quantitative estimates of the direct and indirect radiative forcing is provided by Haywood and Boucher (2000).

d. Explicit treatment of aerosols within GCMs and improved observational capability: 2000s

As computing power increased and advances were made through observations and fundamental developments in the physics of aerosols, GCMs increasingly incorporated detailed aerosol chemistry, transport, and microphysics schemes, frequently on a species-by-species basis under the assumption of external mixing. Models began to incorporate enough of the natural and anthropogenic species prevalent in the atmosphere to make meaningful intercomparisons against observations from, for example, surface-based AERONET sun photometer sites and satellite retrievals. However, determining differences in performance between the models was complicated by the impacts of differing emissions, differing time periods of analysis, and different analyses of observational constraints (Textor et al. 2007). Hence, many modeling centers joined the AeroCom initiative (Kinne et al. 2006; Schulz et al. 2006), which provided a platform for consistent model–model and model–observation intercomparisons. AeroCom initially focused on intercomparison of aerosol optical properties (Kinne et al. 2006) and aerosol direct radiative forcing (Schulz et al. 2006). AeroCom’s remit expanded considerably to include a wide range of intercomparisons such as aerosol indirect effects (Penner et al. 2006; Quaas et al. 2009) along with more specific objectives such as comparisons of model-derived vertical profiles against satellite-borne lidars (Koffi et al. 2012) and in situ black-carbon profiles (Schwarz et al. 2010), enabling refinements of model performance.

The majority of aerosol schemes at this time treated aerosols in “bulk” form; that is, aerosol mass was transported, but the detailed description of aerosol microphysics was not included. Aerosol size distributions and hygroscopic growth factors were assumed based on in situ measurements of aerosol properties from surface sites, in situ aircraft-based measurements, or validated AERONET sky-radiance-based retrievals (Dubovik and King 2000; Haywood et al. 2003). At the time, the impacts of relative humidity on the radiative forcing of aerosols via their influence on optical properties and in particular the specific extinction coefficient were frequently accounted for via measurements of the hygroscopic growth using, for example, airborne humidified nephelometer systems. For example, Kotchenruther and Hobbs (1998) and Kotchenruther et al. (1999) provided hygroscopic growth parameterizations for biomass burning aerosols in Brazil and industrial pollution off the east coast of the United States, respectively.

In addition to aerosol direct effects and aerosol indirect effects, the aerosol semidirect effect started to receive some considerable attention. The semidirect effect is the mechanism whereby aerosol absorption leads to heating of the atmospheric column, increasing atmospheric stability and decreasing the relative humidity. These impacts were postulated to inhibit cloud formation in the layer of absorbing aerosols but also alter cloud cover in other parts of the troposphere (e.g., Ackerman et al. 2000; Johnson et al. 2004; Hansen et al. 1997b). Diagnosing the semidirect effect of aerosols using the strict definition of radiative forcing (holding all other atmospheric variables fixed) was not possible; as for aerosol indirect effects beyond the Twomey effect, this posed a significant problem in quantification of radiative forcing.

MODIS Terra started producing \( \tau_{\text{aer}} \) data in 2002, with MODIS Aqua following in 2002 (Remer et al. 2002); they are still providing essential data for validation and data assimilation to this day. Other satellite sensors provided considerable additional information (e.g., MISR; Kahn et al. 2005; AATSR/AATSR-2; Table 2.2 of Forstner et al. 2007). However, the combination of the near-global daily coverage, developments of retrievals over land surfaces (Hsu et al. 2006), cross calibration with the highly accurate AERONET network, ease of data access, and the longevity of this dataset has resulted in MODIS becoming the mainstay for model validation for both \( \tau_{\text{aer}} \) and for examining aerosol indirect effects via relationships between aerosol and clouds (Quaas et al. 2009). Development of near-global coverage of \( \tau_{\text{aer}} \) from satellites and accumulation mode fraction from the MODIS instrument augmented by in situ aircraft-based measurements allowed the first mainly observational estimate of aerosol direct effects (Bellouin et al. 2005) compared to the earlier efforts.
using model–observation analysis. The anthropogenic fraction was recognized as being almost entirely in the accumulation mode, while natural aerosols in the form of sea salt and mineral dust are typically in the coarse mode, allowing a first observational estimate of the perturbation of $\tau_{\text{aer}}$ by anthropogenic emissions and an associated $F_{\text{direct}}$ of $-0.8 \text{ W m}^{-2}$. Similar methods followed (Chung et al. 2005, $-0.35 \text{ W m}^{-2}$; Yu et al. 2006, $-0.5 \text{ W m}^{-2}$); these estimates were generally rather stronger than those from models, potentially due to the fact that absorbing aerosols (e.g., anthropogenic biomass burning aerosols above clouds) were neglected, which can frequently produce positive radiative forcings (e.g., Keil and Haywood 2003).

More sophisticated estimates of the direct radiative effect of aerosol in cloud-free skies over oceans were also developed by correlating the cloud-free TOA upward solar irradiance (frequently derived from CERES) against the aerosol optical depth derived from other instruments such as VIRS or MODIS (Loeb and Kato 2002; Zhang et al. 2005; Loeb and Manalo-Smith 2005). These estimates provide additional validation data for testing relationships in GCMs but cannot be used to infer the radiative forcing by themselves owing to lack of knowledge of preindustrial conditions.

Aerosol indirect forcing based on satellite retrievals were also developed. Typically, these studies developed relationships between CDNC and fine-mode aerosol concentrations or optical depth (Quaas et al. 2006; Quaas and Boucher 2005). For example, Quaas and Boucher (2005) developed relationships between observed cloud properties from MODIS and observed aerosol properties from POLDER for stratiform marine clouds and for convective clouds over land and utilized these relationships within GCMs. Various methods for partitioning the observed relationships as a function of meteorology, above-cloud moisture, and SSTs were to be developed to account for the impacts of meteorology that can confound derived relationships in observational studies. However, a persistent problem with these correlative studies is the mutual exclusivity of aerosol and cloud satellite retrievals and the lack of account of the relative vertical profile of aerosol and cloud.

IPCC (2007) and Forster et al. (2007) recognized that the growing number of different aerosol species that were being considered in climate models was becoming unwieldy; while aerosol components were still assigned individual radiative forcing values, only the total aerosol direct effect and cloud albedo effect were included on the bar chart. Direct radiative forcing estimates were predominately model based, relying on a combination of AeroCom/non-AeroCom estimates and revealed $F_{\text{direct}} = -0.5 \pm 0.4 \text{ W m}^{-2}$ (5%–95% confidence), while the Twomey effect was estimated to be $-0.7 \text{ W m}^{-2}$ (best estimate) with a 5%–95% range of $-0.3$ to $-1.8 \text{ W m}^{-2}$ (Forster et al. 2007). For the first time, Forster et al. (2007) presented the spread in the long-lived GHG, aerosol, and total radiative forcing using a Monte Carlo simulation of the uncertainties associated with each of the forcing mechanisms (Boucher and Haywood 2001) to demonstrate that the uncertainty in the total radiative forcing was dominated by that of aerosols, particularly owing to uncertainties in the aerosol indirect effects, and that the total radiative forcing was positive, consistent with the observed warming of climate. Forster et al. (2007) recognized that interactions of aerosols with mixed-phase and ice clouds continued to be impossible to quantify on a global-mean basis owing to the even greater complexity of these clouds when compared to warm liquid-phase clouds.

e. Increases in aerosol model complexity—Second-generation models: 2010s

As aerosol modeling matured, further refinements of aerosol direct and indirect effects were included in GCMs. Further components of aerosol were included; Bellouin et al. (2011) included nitrate aerosol and pointed out that as sulfur dioxide emissions decrease in the future owing to emission control, the radiative forcing of nitrate will likely increase owing to the availability of excess ammonia in the atmosphere. However, nitrate continues to remain a difficult aerosol to model owing to the dissociation to nitric acid and ammonia under ambient temperature and humidities.

The development of aerosol mass spectrometers and their location at surface sites and airborne platforms enabled, for the first time, a full appreciation of the complexity of optically active submicron aerosol composition as a function of location and altitude to be deduced (Jimenez et al. 2009) with sulfate, organic, and nitrate highlighted as the dominant submicron components. The problems of the mutual exclusivity of satellite retrievals of aerosol and cloud can be avoided using active satellite sensors such as Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) lidar aerosol data collocated with MODIS cloud data (e.g., Costantino and Bréon 2013). However, in situ airborne platforms with dedicated instrumentation such as nephelometers and aerosol optical particle counters continued to provide vital information on the aerosol vertical profiles at a level of detail and vertical resolution impossible to achieve with satellite-mounted lidars.

In modeling, dual-moment schemes became more common, treating both aerosol number and aerosol mass prognostically (Stier et al. 2005) and both internal
and external mixtures. This has particular relevance to estimates of aerosol–cloud interactions because, for single-moment schemes with prognostic mass only, any increase in the aerosol mass (e.g., via condensation or coagulation) must artificially increase the aerosol number and hence CCN, which then produces stronger aerosol indirect effects. The use of dual-moment state-of-the-art aerosol schemes in GCMs is now commonplace.

Lohmann et al. (2010) examined the differences between (i) the radiative flux perturbation (RFP; Haywood et al. 2009a), which is calculated as the difference in the top-of-the-atmosphere radiation budget between a present-day simulation and a preindustrial simulation, both using the same sea surface temperatures, and (ii) the radiative forcing computed from two calls to the radiative transfer code in GCMs holding the atmospheric state fixed. The RFP calculation allows for rapid responses (e.g., in clouds), which occur on a faster time scale than the large-scale shifts in climate response that are induced through SST responses. RFP has become more commonly known as the ERF [see section 2c(6)]. Allowing rapid adjustment to occur in diagnosing the ERF allowed isolation of both the Twomey (1977) and the Albrecht (1989) aerosol indirect effects and also aerosol semidirect effects.

Myhre (2009) suggested that the discrepancy between observational (stronger) and modeled (weaker) estimates of aerosol direct forcing highlighted in IPCC (2007) was due to (i) the lack of account of aerosol absorption above clouds and (ii) relatively larger fractional increase in BC containing absorbing than scattering aerosols since preindustrial times. Analysis of models reported an aerosol direct radiative forcing of $-0.3 \text{ W m}^{-2}$, which was found to be consistent with observational estimates. Aerosol absorption was again highlighted as a major uncertainty in accurate determination of aerosol direct radiative forcing (Bond et al. 2013) owing to aspects such as the morphology of the black carbon as a function of age and the impact of coatings of organic and inorganic components, the heating in the atmospheric column, and subsequent rapid adjustment. Again this suggested that diagnosing the radiative forcing in a strict sense could not capture the rapid adjustment associated with atmospheric processes.

Boucher et al. (2013), Myhre et al. (2013), and IPCC (2013) recognized that retaining the strict definition of radiative forcing as in previous IPCC reports was becoming untenable because it did not reflect the growing consensus that rapid responses can and should be isolated in any metric of climate change but also because of the ease of application to GCM simulations. Hence, the growth of ERF as the preferred metric for assessing potential climate impacts. Indeed, IPCC also chose the term aerosol–radiation interactions over the aerosol direct forcing and aerosol–cloud interactions over aerosol indirect with rapid adjustment of aerosol–radiation interaction as a term for the semidirect effect (Boucher et al. 2013). By this time, there were many mature estimates of the impact of aerosol–radiation interactions and aerosol–cloud interactions from sophisticated GCMs and satellite-based estimates (e.g., Fig. 14-17; see also Tables 7.4 and 7.5 of Boucher et al. 2013), allowing Myhre et al. (2013) to estimate the magnitude of preindustrial to present-day aerosol–radiation interactions ($-0.45 \text{ W m}^{-2}$ with a 95% uncertainty range of $-0.9$ to $+0.05 \text{ W m}^{-2}$) and aerosol–cloud interactions ($-0.45 \text{ W m}^{-2}$ with a 95% uncertainty range of $-1.2$ to $0 \text{ W m}^{-2}$).

**f. Current promising lines of research**

We have seen that the radiative forcing (or ERF) has changed significantly from best estimates of stronger than $-2 \text{ W m}^{-2}$ (Charlson et al. 1991, 1992) to weaker than $-1 \text{ W m}^{-2}$. Much of this reduction in magnitude of the radiative forcing via the direct effect (aerosol–radiation interactions) was captured by the late 1990s owing to the use of GCMs (e.g., Kiehl and Briegleb 1993) and by accounting for the effects of aerosol absorption by black carbon (Haywood and Shine 1995). However, uncertainties in both the aerosol–radiation and aerosol–cloud interactions have remained stubbornly difficult to reduce owing to structural and parametric uncertainties. Here, we ask what progress has been made in the five years since IPCC (2013) in terms of promising avenues of research.

The direct radiative effect/forcing of aerosols above clouds has remained a contentious issue with some very strong instantaneous positive radiative effects (greater than $+130$ to $+150 \text{ W m}^{-2}$) being diagnosed from various satellite instruments (e.g., de Graaf et al. 2012; Meyer et al. 2015; Peers et al. 2015) over the southeast Atlantic, values that are stronger than those from climate models (de Graaf et al. 2014). Zuidema et al. (2016) highlights that global models diverge when determining the direct radiative effect in the region. This is because the direct radiative forcing of a partially absorbing aerosol such as biomass burning aerosol depends not just on determination of the aerosol optical depth and aerosol absorption properties but on the cloud amount, cloud reflectance, and the relative vertical profile of cloud relative to the aerosols (see Fig. 14-18). The ORACLES, LASIC, CLARIFY, and Aerosol Radiation and Clouds in Southern Africa (AeroClo-SA) in situ aircraft-based measurement campaigns have targeted deriving better estimates of the direct effect of
absorbing aerosols over clouds as one of their primary objectives (Zuidema et al. 2016). These measurement campaigns will undoubtedly give a better understanding of direct radiative effects of partially absorbing aerosols above clouds in the key region of the southeast Atlantic, but the change in concentration from preindustrial times may well preclude accurate determination of the radiative forcing.

Accurate representation of preindustrial aerosol concentrations is also highlighted as a key uncertainty by Carslaw et al. (2013), who used a statistical emulator approach to examine the sensitivity of the aerosol forcing to a wide range of parameters including uncertainties in anthropogenic and natural emissions. Uncertainties in natural aerosol concentrations accounted for around 45% of the variance, while uncertainty in anthropogenic emissions accounted for around 35% of the variance due to the impact that natural background aerosols have on the susceptibility of clouds to anthropogenic perturbations. Based on the
limited paleo data, desert dust may have increased by almost 40% over the twentieth century due to a combination of climate change and land use (Mahowald et al. 2010; Ginoux et al. 2012). An important and likely source of higher preindustrial aerosols, which is currently poorly constrained, are wildfires. Recent studies have suggested that, due to uncertainties in preindustrial wildfire emissions alone, the range of anthropogenic indirect effects could be between $-0.1$ to $-1.1$ W m$^{-2}$ (Hamilton et al. 2018). If these estimates are supported by more studies, we would have to rethink how we constrain estimates of both aerosol indirect effects from anthropogenic aerosols and climate sensitivity (see section 6 for more discussion).

Progress been made in understanding the impact of anthropogenic emissions on mixed-phase cloud by contrasting observations of their behavior against liquid-phase clouds. Christensen et al. (2014) examined ship track data in mixed-cloud environments and found a more muted indirect radiative forcing impact owing to enhanced glaciation-induced precipitation that limited the total water path of the clouds. Christensen et al. (2016) extended these observations by using multiple sensors over millions of atmospheric profiles and concluded that liquid clouds dominate any negative radiative forcing for the aerosol indirect effects owing to the muted impacts of mixed-phase clouds and a counterbalancing positive radiative forcing from convective clouds. These observations call into question whether net aerosol indirect effects have been overestimated.

Ghan et al. (2016) performed an intercomparison to isolate the strength of the Twomey and Albrecht aerosol–cloud-interaction effects in GCMs and showed that while all models exhibited a reasonably consistent Twomey effect, the strength of the Albrecht effect essentially fell into two clusters. The clusters were (i) an almost negligible impact and (ii) a strong positive forcing that acted to reinforce the radiative forcing from the Twomey effect, but observational evidence remained
lacking as to which one of these responses was correct. However, one interesting line of evidence to elucidate
the strength of the Albrecht effect was the use of large-
scale SO$_2$-degassing volcanic eruptions in relatively
pristine environments to examine the impact on satellite-
derived cloud properties. This technique was first used by
Gassó (2008), on relatively modest degassing events and
has been the subject of further research (Toll et al. 2017).
These smaller-scale degassing eruptions can be used to
examine relationships between cloud and aerosol in a
similar way to ship tracks (Christensen et al. 2014; Chen
et al. 2015) but are frequently on too small a scale for
the impacts to be directly compared against GCMs ca-
pable of diagnosing a radiative forcing. This situation
changed in 2014 with the fissure eruption at Holuhraun,
Iceland, which emitted a huge plume of SO$_2$ across the
entire North Atlantic (Gettelman et al. 2015) causing a
clear, statistically significant reduction in the cloud ef-
fective radius in the MODIS satellite record (McCoy and
Hartmann 2015; Malavelle et al. 2017) but no discernible
impact of the total liquid water path. Malavelle et al.
(2017) were able to show that only models with a modest
Albrecht effect were consistent with observations for a
host of liquid-water cloud conditions. In a similar vein,
Chen et al. (2014) showed that variations in cloud liquid
water paths are dominated by other meteorological fac-
tors such as the state of precipitation, humidity, and at-
mospheric stability rather than aerosol microphysical
processes making definitive detection and attribution
difficult.

These advances provide an opportunity to better
constrain aerosol indirect effects in the future; without
pursuing these opportunities to confront model per-
formance, an accurate characterization of aerosol ra-
diative forcing is likely to prove elusive. There is,
however, a growing consensus that, in the future, the
importance of the radiative forcing of aerosols will
begin to be a less important uncertainty as the radiative
forcing from greenhouse gases continues to increase.
Global emissions of sulfur dioxide have plateaued at
around 1990 and have begun to fall on a global-mean
basis owing to effective clean-air policies targeted at
reducing particulate pollution. Areas such as the
United States and Europe have already seen reduct-
tions in sulfur dioxide emissions of around a factor of 5
since their peak. This reduction in emissions, coupled
with the ever increasing radiative forcing from green-
house gases may result in significant rates of global
warming over the next few decades (e.g., Andere et al. 2005).

Where aerosols may start to play an increasing role is
in the, currently theoretical, field of geoengineering, that
is, the deliberate injection of aerosols or their precursors
into the stratosphere to mimic the cooling impacts of
explosive volcanic eruptions such as Pinatubo or the
deliberate injection of aerosols into stratocumulus
clouds to mimic the impacts of natural degassing vol-
canoes. The use of such techniques would have many,
many consequences (e.g., Robock et al. 2008b), not least
that, if proven effective, it could reduce the drive for
reduced use of fossil fuels (see section 13 for a com-
prehensive discussion).

6. Land and biogeochemistry interactions

Human activities do not only directly emit gases and
aerosols that impact climate, as described above, they
also modify the land surface, which can directly change
the surface properties, and both directly and indirectly
change the emissions of different gases and aerosols
(Feddema et al. 2005; Heald and Spracken 2015; Myhre
et al. 2013; Pielke 2005; Ward et al. 2014). Land con-
version of forests, for example, to croplands, emits car-
bon dioxide immediately, and the land-cover change
and the land management has many implications for
albedo changes and emissions. Changes in land surface,
such as urbanization or deforestation, can also change
the local experience of climate in substantial ways, but
these are not contributing to changes in the top-of-
the atmosphere radiative forcing directly and thus are not
discussed here (Field et al. 2014; Hartmann et al. 2013;
Lejeune et al. 2018). We also do not consider longer-
term feedbacks, such as the fertilization of land or ocean
ecosystems by anthropogenic aerosols, which could be
as large as the direct radiative effects of anthropogenic
aerosols (Mahowald 2011; Mahowald et al. 2017a). Gener-
ally speaking, understanding and quantifying how
changes to the land surface are impacting the climate
are more difficult than some of the direct emissions
from human activities described previously (Boucher

The most important impact of human land-use and
land-cover change (LULCC), in terms of radiative
forcing of climate, and one of the first processes included
in the IPCC assessment reports are the direct emis-
sions of CO$_2$ from deforestation and the indirect
emissions of CO$_2$ from forest degradation and land
management (IPCC 1990). Other processes, such as
the albedo changes, interactions with wildfires, and
other processes, were mentioned in the first report but
not quantified (IPCC 1990). In subsequent AR, these
processes were expanded upon and quantified, as
described below.

For long-lived gases, which can be assumed to be
relatively well mixed across the troposphere, such as
carbon dioxide, methane, or nitrous oxides, estimates
of changes in the atmospheric composition rely on ice core and other observational records, as discussed previously (Forster et al. 2007; Myhre et al. 2013). But for short-lived gases and aerosols, an important difficulty in understanding how human activities have changed past atmospheric conditions, especially, for example, in understanding baseline preindustrial chemistry or aerosol conditions, is that there are no direct observations of preindustrial conditions for short-lived gases or aerosols, since the ice core records from just a few sites cannot characterize the global average for short-lived constituents (Myhre et al. 2013). For short-lived gases and aerosols, their emissions are estimated based on simple assumptions or modeling studies (Boucher et al. 2013; Collins et al. 2017; Lamarque et al. 2010; Van Marle et al. 2017; Myhre et al. 2013; Shindell et al. 2013b).

Recent studies have highlighted the importance of potential changes in emissions from “natural” sources due to human activities, especially on the land surface (Carslaw et al. 2010, 2013; Hamilton et al. 2018; Mahowald et al. 2011, 2010; Myhre et al. 2013; Ward et al. 2014), so here, we review the state of knowledge on how humans can perturb natural sources. We will also review the evolution of the representation of human land-use forcings of climate in the IPCC reports, as an indication of how the understanding of how human activities on the land surface impacting natural sources of different important radiative constituents has evolved.

### a. Land-use and land-cover change climate forcers

The interactions of land-use and land-cover change are shown in Fig. 14-19 and described in more detail in each section below.

#### 1) ALBEDO IMPACTS FROM LAND-USE AND LAND-COVER CHANGE

The direct modification of the surface from land-use, land-management, or land-cover change can be through albedo changes or through changes in the energy fluxes locally (Andrews et al. 2017; Bonan 2008; Feddema et al. 2005; Myhre et al. 2013; Pielke 2005). The surface albedo (ratio of the reflected and incoming solar radiation) varies between dark forests, lighter grasslands or crops, and often even lighter barren deserts or snow-covered surfaces. Consequently, deforestation tends to increase Earth’s albedo and causes more light to be reflected, cooling Earth. The largest impacts of land use and deforestation will occur at higher latitudes because snow-covered forests retain a low albedo, while snow-covered croplands will have a high albedo (IPCC 2001). Recently, satellite measurement have been used to estimate the impact of land use on radiative forcing (Myhre et al. 2013; Zhao and Jackson 2014). Estimates of the changes since preindustrial require the use of models,
and the biophysical impacts of LULCC are sensitive to the model’s used; a land-use model intercomparison project (LUMIP) is underway to better constrain the impacts on radiative forcing (Lawrence et al. 2016). Changes in the land surface, such as changes in vegetation, and their impact on the surface albedo have been included in the assessments since the first report (IPCC 1990), but not quantified until the third report, to be $-0.2 \pm 0.2 \text{ W m}^{-2}$ (IPCC 2001). The estimates and uncertainty stayed the same in the AR4 but became slightly smaller, with smaller uncertainties in the AR5 ($-0.15 \pm 010 \text{ W m}^{-2}$) (Myhre et al. 2013).

2) CARBON DIOXIDE EMISSIONS

The gross carbon dioxide emissions or uptake from natural ecosystems, both land and ocean, are much larger than the fluxes from anthropogenic emissions, but the net emissions are close to zero (Ciais et al. 2013b; Le Quéré et al. 2013, 2016; Sitch et al. 2015). Deforestation and conversion of natural lands for human management tend to directly release carbon dioxide into the atmosphere, while reforestation, afforestation, or carbon dioxide fertilization will increase the uptake of carbon dioxide into the terrestrial system, including soils (Ciais et al. 2013b; Friedlingstein et al. 2006; Houghton 2018; Le Quéré et al. 2013). In addition, the change in land cover and land management changes the longer-term uptake of carbon dioxide (Ciais et al. 2013b; Friedlingstein et al. 2006; Houghton 2018; Le Quéré et al. 2013). Over all of the IPCC assessment reports, ice core data are used for carbon dioxide changes, which do not require the attribution of the change in carbon dioxide from land surface changes versus direct emission (Ciais et al. 2013b; Denman et al. 2007; IPCC 1996, 2001; IPCC 2013). However, starting already in the FAR, there was an important division between land-use emissions and other emissions identified in the reports (IPCC 1990). The 2000 special report on land use, land-use change, and forestry was written in support of the Kyoto Protocol and provided guidance on how to calculate carbon in land conversion.

3) WILDFIRES AND BIOMASS BURNING

Natural ecosystems can emit short-lived gases and aerosols through a variety of processes, and these processes can be modified by human activities (Carslaw et al. 2010, 2013; Hamilton et al. 2018; Mahowald et al. 2011, 2010; Myhre et al. 2013; Ward et al. 2014) (Fig. 14-19). Of the most important is likely to be changes in wildfire regimes (Arneth et al. 2010; Carslaw et al. 2010; Ward et al. 2014). Intermittent wildfires release substantial amounts of carbon dioxide, carbon monoxide, sulfur dioxide, methane and other organic gases, and carbonaceous aerosols (Andreae and Merlet 2001). Wildfire sources of methane were identified already in the FAR. Biomass burning aerosols were one of the first three aerosols considered in the TAR (along with industrial sources of sulfate and black carbon). For AR4 and AR5, studies suggested that preindustrial wildfires were much less than in current climate, as they attributed wildfires to deforestation fires (Boucher et al. 2013; Forster et al. 2007; Lamarque et al. 2010). Recent evidence suggests strong changes in wildfire over the last 250 years, with a large increase in fires about 200 years ago and then a decrease (Kloster et al. 2010; Marlon et al. 2008, 2012; Pechony and Shindell 2010; Zennaro et al. 2014). Recently, evidence from satellites has suggested a 25% decrease in fires over the last 18 years, likely due to the expansion of agriculture (Andela et al. 2017). While increases in biomass burning due to deforestation and climate change are assumed to have occurred since preindustrial times in the standard estimates used for CMIP6 (Van Marle et al. 2017), some estimates suggest a decline instead, due largely to a change in human fire and land management (Hamilton et al. 2018; Marlon et al. 2008; Zennaro et al. 2014). These large decreases in wildfire emissions over the Anthropocene are large enough to have an impact on indirect and direct aerosol radiative forcing so that they offset those from direct anthropogenic emissions (Hamilton et al. 2018). For example, that study suggested that new wildfire emission estimates increase estimated anthropogenic aerosol indirect forcing from $-1.1 \text{ W m}^{-2}$ using the CMIP6 emission datasets to $0.1 \text{ W m}^{-2}$ using LMfire estimates of wildfire (Fig. 14-20; using global averages from Hamilton et al. 2018). In other words, because the preindustrial wildfire
emissions are poorly constrained, current-day anthropogenic aerosol radiative forcing could be \(-1.1\) or \(0.1 \text{ W m}^{-2}\) using the same model but different emission estimates. This has profound implications for our understanding, not only of anthropogenic aerosol radiative forcing but also climate sensitivity, which is very sensitive to assumptions about the size of the anthropogenic aerosol radiative forcing (Knutti et al. 2002; Myhre et al. 2013).

4) **Agricultural Activity and Soils**

Soils naturally release nitrogen oxides, nitrous oxide, and ammonia: changes in nitrogen inputs or temperature can radically modify the amount of nitrogen oxides or ammonia released (Ciais et al. 2013b; Fowler et al. 2013). Since agriculture, and especially the green revolution, there has been a substantial modification of the nitrogen budgets of regions with land use, modifying substantially the nitrogen inputs (Ciais et al. 2013b; Fowler et al. 2013). In terms of direct radiative forcing of nitrogen-based species, nitrous oxide is the most important and is a long-lived greenhouse gas, with a lifetime on the order of 100 years (Ciais et al. 2013b). Already in the FAR, the agricultural sources of nitrous oxide were identified, if not quantified (IPCC 1990), while in the SAR and later reports, estimates for the agricultural sources of nitrous oxides were quantified (Ciais et al. 2013b; Denman et al. 2007; Forster et al. 2007; IPCC 1996, 2001; Myhre et al. 2013). The ice core changes over the last 100 years in nitrous oxide have been attributed to changes in land management (Ciais et al. 2013a; Ward et al. 2014). Nitrogen oxide and ammonia emissions from soils are thought to be enhanced by agricultural activities, especially nitrogen fertilizers (Fowler et al. 2013; Myhre et al. 2013), and the role of human land use in modifying these emissions from soils is first mentioned in the TAR (IPCC 2001). Nitrogen oxide emissions are important climatically for their impact on tropospheric ozone and methane lifetime (section 4), but the anthropogenic part is dominated by combustion sources (Myhre et al. 2013; Shindell et al. 2017). On the other hand, ammonia emissions are predominantly from nitrogen fertilization as part of agriculture or pasture usage, which contribute to a change in ammonium aerosols (Myhre et al. 2013; Riddick et al. 2016; Shindell et al. 2017; Sutton et al. 2013).

5) **Methane**

In inundated regions, methanogens thrive, producing methane from organic material in the soil (Matthews and Fung 1987). Changes in the area of inundated areas (e.g., due to expansion of rice paddies or filling of wetlands), productivity of these regions, temperatures, and carbon dioxide itself have impacted the methane emissions from these regions (Kirschke et al. 2013; Myhre et al. 2013; Paudel et al. 2016; Zhang et al. 2017). In addition to changes in natural wetland area, rice paddy expansion and ruminant animal husbandry has increased land-use methane production (Kirschke et al. 2013; Myhre et al. 2013). Emissions of methane from land-use change were already considered early in the IPCC process (IPCC 1990).

6) **Mineral Aerosols (Dust)**

Dry, unvegetated land subject to strong winds allow the entrainment of soils into the atmosphere, causing the largest source of aerosols by mass into the atmosphere (Mahowald et al. 2011). Desert dust is one of the few aerosols for which we have paleo records, since it can be retrieved from ice, marine, terrestrial, and land records (Albani et al. 2018; Kohfeld and Harrison 2001; Mahowald et al. 2010). AR4 assumed that <10% of mineral aerosols were from anthropogenic sources (Forster et al. 2007). A reconstruction based on paleo data suggests that dust may have increased by a factor of almost 2 across the twentieth century (Mahowald et al. 2010), due to either aridification from climate change (Mahowald 2007) or land use (Ginoux et al. 2012). However, between the AR4 and AR5, estimates of the radiative forcing from mineral aerosols became closer to zero because of a shift in dust properties (Albani et al. 2014; Mahowald et al. 2010; Perlwitz et al. 2001; Sinyuk et al. 2003), suggesting small contributions from anthropogenic desert dust to radiative forcing. Mineral aerosols absorb and scatter in both the shortwave and longwave, making them complicated in their impacts (Sokolik and Toon 1996). The shift to a smaller-magnitude radiative forcing for mineral aerosols is due to both an improved estimate that mineral aerosols are likely more absorbing in the shortwave than previously thought from remote sensing data (Sinyuk et al. 2003), as well as a consensus that mineral aerosols tends to be larger than previously thought (Kok et al. 2017; Mahowald et al. 2014). Future projections vary depending on whether they include climate change impacts (Evan et al. 2016), land-use impacts, or both (Ward et al. 2014). While the SAR included the possibility of anthropogenic land-use sources of desert dust (IPCC 2001; Tegen et al. 1996), the radiative forcing of this constituent was not included in the report until the TAR (Forster et al. 2007).

7) **Organic Compounds**

Some plants emit volatile organic compounds (Guenther et al. 2006), which can interact with nitrogen oxides to change the cycling of tropospheric ozone...
of LULCC on radiative forcing (Forster et al. 2007). The AR5 estimate is slightly smaller (Collins et al. 2017; Myhre et al. 2013), as well as produce secondary organic aerosols from fungi, pollen, or plant or insect pieces (Despres et al. 2012; Graham et al. 2003; Mahowald et al. 2011). Deforestation, climate change, or changes in fire frequency can modify the amount of forests, thereby modifying the emissions from natural forests of these important constituents (Arnot et al. 2010; Mahowald et al. 2011; Unger 2014; Ward et al. 2014). The importance of biogenic-derived organics for modifying both ozone and secondary organic aerosol formation was established by the TAR, although no explicit calculation of the impact of human land use or fires onto the radiative forcing was performed until after the last assessment (Forster et al. 2007; IPCC 2001; Myhre et al. 2013) and is thought to be a small, but important feedback (Unger 2014; Ward et al. 2014).

b. Snow albedo changes

Anthropogenic aerosol deposition of black carbon onto snow can change the albedo of the snow, darkening the snow and warming the globe (Hansen and Nazarenko 2004) as well as modifying the melting of the snow and glaciers (Painter et al. 2013). This effect was first included in the AR4 as a slight warming: +0.1 (0.0–0.2) W m$^{-2}$ (Forster et al. 2007). The AR5 estimate is slightly smaller at +0.04 (0.02–0.09) W m$^{-2}$.

c. Estimates of the net effect of LULCC on radiative forcing

To better understand the relative role of LULCC compared to other sources of radiative forcing, one set of studies split the emissions into those from LULCC and non-LULCC based on standard CMIP5 and IPCC AR5 input datasets and approaches, including using the ERF concept (Mahowald et al. 2017b; Ward et al. 2014; Ward and Mahowald 2015). Similar to most climate models, the model used in this set of studies overpredicted aerosol direct and indirect effects compared to IPCC AR5 assessed RF, and thus, the aerosol radiative forcings from LULCC results were tuned to the AR5 estimates (Mahowald et al. 2017b; Ward et al. 2014; Ward and Mahowald 2015). Radiative forcing from the LULCC sector represents 40% of the current anthropogenic forcing (Fig. 14-21a). This is due to the carbon dioxide emissions from conversion of natural lands to managed lands, in addition to substantial radiative forcing from methane and nitrous oxide emitted from agriculture and changes in land (Mahowald et al. 2017b; Ward et al. 2014; Ward and Mahowald 2015). Anthropogenic aerosol changes from land use are thought to be large individually, due to changes in desert dust, agricultural aerosols, forest biogenic aerosol emissions, and wildfires but have a net zero impact on radiative forcing in this set of studies (Ward et al. 2014) so that most of anthropogenic aerosol radiative forcing is due to non-LULCC (Fig. 14-21a), which is a net negative radiative forcing. The fraction of the total radiative forcing from LULCC (40%) is larger than the fraction of the CO$_2$ radiative forcing attributable to LULCC (20%). Over time, the LULCC radiative forcing has grown gradually over the twentieth century, while non-LULCC radiative forcing was close to zero until the 1970s and now is growing very quickly to positive values as CO$_2$ is accumulating in the atmosphere (Fig. 14-21b).

d. Future projections

LULCC RF for future are even more difficult to estimate than that for past and are highly dependent on driving assumptions, especially how much land-use conversion will occur (Ward et al. 2014). Integrated assessment models that were used to create the forcing scenarios for Earth system models (Gidden et al. 2019; Hurtt et al. 2011; Moss et al. 2010) tended to underestimate current deforestation rates in the AR5, especially in the tropics, and tended to have very similar deforestation rates, compared with possible futures (Ciais et al. 2013b; Ward et al. 2014). This suggests that future estimates from integrated assessment models (IAMs) or IAMs coupled to Earth system models may underestimate the impact of LULCC (Mahowald et al. 2017b). The LUMIP has the goal to explore more fully the possible LULCC pathways as well as the radiative forcing resulting from the land biophysics component (Lawrence et al. 2016).

The emission datasets used in the CMIP5 and CMIP6 include some of the processes that might impact radiative forcing from LULCC: generally only direct emissions from agriculture are included in emission changes, including some estimates of changes in wildfires in the past and future but with no changes in desert dust (Collins et al. 2017; Gidden et al. 2019; Lamarque et al. 2010, 2011). However, the CMIP6 studies, currently underway, will include idealized sensitivity studies for the different natural aerosols to understand their impact on current radiative forcing, providing some bounding on their current and future role (Collins et al. 2017).

7. Contrails and contrail-induced cloudiness

One important component to the RF from subsonic and supersonic aircraft operations (in addition to its direct emissions of CO$_2$ and the indirect effects of NO$_x$ emissions) arises from the ejection of hot moist air (and aerosols and their precursors) from jet engines into air
that is supersaturated (with respect to ice) in the upper troposphere and lower stratosphere (e.g., Heymsfield et al. 2010; Kärcher 2018). This leads to the formation of persistent linear condensation trails (contrails) that can evolve into persistent cirrus (e.g., Minnis et al. 1998; Boucher 1999; Fahey et al. 1999), composed mostly of ice condensates. Contrail characteristics are influenced by humidity and temperature along aircraft flight tracks. In addition, aerosols from aviation can alter the properties of existing clouds or influence their subsequent formation. The potential impact of aviation contrails on climate was recognized in the early 1970s. Machta and Carpenter (1971) presented evidence of changes in high cloud cover over Denver, Colorado, and Salt Lake City, Utah, which they tentatively attributed to aviation; in the same volume, Matthews et al. (1971) noted that, while this would increase the albedo, a compensating change in longwave emissivity could reverse the sign of the climate effect.

Assessments of contrail RF have mostly concluded that, when diurnally averaged, the net effect of contrails is positive, due to the dominance of the longwave component. An early quantification of persistent linear contrails consisted of using meteorological and air traffic data scaled to regional observations of contrail cover (Minnis et al. 1999; Fahey et al. 1999). These studies yielded an estimate of 0.02 W m$^{-2}$, with Fahey et al. (1999) estimating a range: 0.005 to 0.06 W m$^{-2}$ (IPCC 1999, 2001). The uncertainty factors included contrail cover, optical depth, and cloud microphysical properties, which runs true even today. Sausen et al. (2005) updated the IPCC estimate, giving a year 2000 forcing of 0.01 W m$^{-2}$ for persistent linear contrails and a total (including the impact on cirrus) of 0.05 W m$^{-2}$.

Soot emissions from aircraft can also affect the cirrus cloud processes, especially the nucleation of ice crystals, but the effects are deemed highly uncertain. Sedimentation of ice crystals from contrails may remove water vapor from the upper troposphere with resultant impacts on vertical profiles of cloud condensates and humidity.

![Fig. 14-21. (a) Anthropogenic radiative forcing (W m$^{-2}$) for the year 2010 relative to 1850 partitioned into LULCC and non-LULCC sources for different forcing agents, with uncertainty in the portion of total radiative forcing due to LULCC given by the error bars [adapted from Ward et al. (2014)]. (b) Time series of anthropogenic radiative forcing (W m$^{-2}$) for LULCC and non-LULCC sources with uncertainty represented by the shading. [From Mahowald et al. (2017b), copyright 2017, Mahowald et al.; https://creativecommons.org/licenses/by/3.0/].](image-url)
Boucher et al. (2013) used more advanced model results to support the view that the longwave forcing dominates over the shortwave (e.g., Stuber and Forster 2007); however, models disagree on the relative importance of the two effects. Contrails have been observed to spread into cirrus sheets that can persist for several hours, and observations confirm their overall positive RF (Haywood et al. 2009b). Boucher et al. (2013) gave a global-mean RF estimate of 0.01 (0.005 to 0.03) W m\(^{-2}\) for persistent linear contrails. Based on Schumann and Graf (2013) and Burkhardt and Kärcher (2011), the combined linear contrails and contrail-induced cirrus ERF for the year 2011 was assessed to be 0.05 (0.02 to 0.15) W m\(^{-2}\), with the principal uncertainties still occurring due to gaps in the knowledge of the spreading rate, physical properties including optical depth and shape, radiative transfer, as well as the lack of knowledge of actual aircraft trajectories. This emphasizes the importance of ongoing measurement campaigns that target contrail properties (e.g., Voigt et al. 2017) as well as the systematic collation of existing knowledge emerging from such campaigns (Schumann et al. 2017). Analysis of satellite data, especially in the context of detecting line-shaped contrails, remains important in the assessment of contrail RF. The recent estimate of Duda et al. (2019), using MODIS data for the period 2006–12, yielded global-mean RF of 0.008–0.009 W m\(^{-2}\), supporting the values assessed earlier in Boucher et al. (2013).

Bock and Burkhardt (2016) have produced a more-refined model estimate of contrail-induced cirrus (including line-shaped contrails) forcing of 0.056 W m\(^{-2}\) for a 2006 aviation inventory. Although this seemed in good agreement with the earlier Burkhardt and Kärcher (2011) results, it arose from a different balance between the longwave and shortwave RF components.

Both Chen and Gettelman (2016) and Bock and Burkhardt (2019) have performed detailed modeling studies of the possible contrail RF in 2050. There are formidable challenges in making such estimates, in addition to those that are relevant to estimating present-day forcing. These include assumptions on the growth in air traffic, and the regional distribution of that growth; changes in engine technology and fuel type; and how changes in climate may impact the occurrence of ice-supersaturated regions in which contrails form. Chen and Gettelman (2016) predict a sevenfold increase in contrail forcing from 2006 to 2050, which is substantially more than the fourfold increase in global-flight distance—they attribute this to stronger growth in flight distances in low latitudes. Bock and Burkhardt (2019) obtain a more-modest factor of 3 increase in global-mean forcing and find that growth in air traffic is the predominant cause, with other effects only impacting the regional distribution. The absolute 2050 forcing differs significantly between these studies, with Chen and Gettelman (2016) obtaining a value of 87 mW m\(^{-2}\), while Bock and Burkhardt (2019) obtain 160 to 180 mW m\(^{-2}\). Bock and Burkhardt (2019) attribute the differences to different assumptions of ice crystal size in newly formed contrails; additional effects include differences in the simulation of changes in the distribution of ice-supersaturated regions in a future climate in the two models.

A significant issue in understanding the importance of contrail RF arises from uncertainty in the efficacy of that forcing [see section 2c(4)]. There is a sparse literature on the topic. In GCM studies, Ponater et al. (2006) found an efficacy of 0.6, while Rap et al. (2010) obtained a value of 0.3. Both studies predate the uptake of ERF as a concept, and the reason(s) for these low values remains unclear; however, if confirmed by later work, it would imply a significantly lower impact of contrails on surface temperature than implied by the radiative forcing. Schumann and Mayer (2017), using results from a simple global-mean model, speculate that the climate sensitivity to the (negative) shortwave contrail forcing may exceed that due to the (positive) longwave forcing, because of differences in the partitioning of the surface and top-of-atmosphere forcings in the two cases. As they note, there is a need to test this hypothesis in a comprehensive global model.

8. Solar radiative forcing

Three primary foci compose contemporary solar radiative forcing research: 1) space-based measurements of solar irradiance, which explicitly define the forcing over the past 40 years; 2) modeling observed irradiance variability in terms of proxies that expand understanding across wavelengths and to multicentennial time scales; and 3) detection and understanding of terrestrial responses to solar irradiance variability for indirect assessment of the forcing. Following an historical overview, this section addresses the development and current status of these three primary topics, concluding with a summary of successes, uncertainties, and challenges.

a. Historical overview

Solar photons at all wavelengths of the electromagnetic spectrum interact with the terrestrial system over a range of altitudes, via multiple processes that depend on the composition of Earth’s land, atmosphere, ice, and ocean. The processes couple radiatively, chemically, and dynamically to distribute incoming solar energy from
low to high latitudes and among different altitude regimes. Solar radiation is Earth’s primary energy source, and pursuit of possible terrestrial influences of its variability has a long history. The discovery in 1843 of an 11-yr cycle in the occurrence of dark sunspots on the sun’s surface initiated efforts to detect, understand, and specify variability in solar radiative output and Earth’s response that continue today (see, e.g., reviews by Hoyt and Schatten 1997; Gray et al. 2010; Lean 2017).

From the mid-nineteenth century until the late twentieth century, correlations between solar indices such as sunspots and climate indices such as temperature afforded the primary evidence in support of a solar influence on climate. However, the phase and magnitude of correlations from individual sites sometimes differed from each other and during different epochs, making their significance difficult to establish. Moreover, the sun’s total irradiance was assumed to be invariant; solar-related changes at the level of 0.1% detected in a few decades of ground-based observations made in the early to mid-twentieth century were attributed to changes in atmospheric transmission and speculated to be due to changes in ozone concentrations in response to the sun’s more variable ultraviolet radiation (Foukal Mack and Vernazza 1977). Sporadic balloon- and rocket-borne measurements did detect variations in solar ultraviolet irradiance but with large uncertainties. Exploratory studies of the terrestrial impacts of solar variability using physical models of climate and, independently, of ozone suggested that an increase in total solar irradiance of 2% was equivalent to doubling CO₂ concentrations (Wetherald and Manabe 1975) and that a 30% increase in solar UV irradiance would increase total ozone by 5% (Penner and Chang 1978).

Space-based observations of the sun transformed knowledge of solar radiative forcing. They quantify unequivocal changes in the sun’s total irradiance over the past 40 years (Fröhlich and Lean 2004; Dudok de Wit et al. 2017) and characterize concurrent spectral irradiance changes at all wavelengths (Haberreiter et al. 2017). Models successfully simulate the observed irradiance changes in terms of the occurrence of dark sunspots and bright faculae on the solar disc (Fröhlich and Lean 2004; Lean et al. 2005) and estimate solar radiative forcing during past millennia using solar activity proxies (Jungclaus et al. 2017; Matthes et al. 2017; Lean 2018a). In 2015, NOAA implemented a Solar Irradiance Climate Data Record (CDR; Coddington et al. 2016) in support of multiple solar–terrestrial endeavors.

Space-based and expanded ground-based observations of Earth’s surface and atmosphere provide compelling terrestrial evidence for solar radiative forcing (Gray et al. 2010; Liljenstern et al. 2015; Lean 2017). Solar signals are affirmed in surface and atmospheric temperature and ozone concentrations over multiple recent decades, providing a new framework for verification and interpretation of site-specific sun–climate paleo connections evident in multiple paleoclimate records, especially of hydrological variables (e.g., Haug et al. 2003; Antico and Torres 2015). State-of-the-art physical general circulation climate models, such as those used in recent IPCC assessments, couple surface, atmosphere, and ocean processes, include interactive ozone, and input spectral—not just total—irradiance. They indicate detectable responses to solar radiative forcing throughout the ocean, surface, and atmosphere (Mitchell et al. 2015; Misios et al. 2015; Hood et al. 2015), as do statistical models of the direct observations (Lean and Rind 2008; Lean 2017; Foster and Rahmstorf 2011).

b. Space-based observations of solar irradiance variability

The launch into space of electrical substitution “active” cavity radiometers enabled the detection of real changes in total solar irradiance, beginning with the Hickey–Friedan radiometer on the Nimbus-7 satellite in November 1978 (Hickey et al. 1980). The radiometers compare solar radiant heating of a black cavity of known aperture area with equivalent electrical power. The Active Cavity Radiometer Irradiance Monitor (ACRIM) on the Solar Maximum Mission (SMM; 1980–89), Upper Atmosphere Research Satellite (UARS; 1992–2005), and ACRIMSat (1999–2013) comprised multiple cavities each with different solar exposure (Willson 1979, 2014), thereby isolating real solar irradiance changes during the 11-yr solar cycle from changes in radiometer sensitivity. Radiometers with different configurations of cavity geometry, surface coating, baffles, and aperture placement continue to measure total solar irradiance on the Solar and Heliospheric Observatory (SOHO) since 1996 (Fröhlich et al. 1995; Fröhlich 2013), the Solar Radiation and Climate Experiment (SORCE) since 2003 (Kopp and Lawrence 2005; Kopp and Lean 2011), and on the International Space Station (ISS) since 2018 (Richard et al. 2011). By virtue of advanced radiometric design and signal detection, etched cavity surface, extensive characterization, and ground-based absolute calibration, the Total and Spectral Solar Irradiance Sensor (TSIS) on the ISS is expected to measure total solar irradiance with <100 ppm uncertainty and 10 ppm yr⁻¹ repeatability (Richard et al. 2011; Pilewski et al. 2018).

Space-based observations show unequivocal total solar irradiance variability on time scales from days to decades as the sun rotates on its axis (every 27 days) and in concert with the 11-yr activity cycle. Initial ACRIM observations readily detected the day-to-day variations
(Willson et al. 1981; Hudson et al. 1982), with solar cycle changes subsequently established in the longer ACRIM dataset (Foukal and Lean 1988). Total solar irradiance can decrease by a few tenths of a percent over a few days, but the sun is brighter overall when solar activity is higher, as indicated by higher sunspot numbers.

Figure 14-22 shows monthly mean values of sunspot numbers and total solar irradiance since 1978 according to five different composite records constructed by cross calibrating and combining multiple observations. The absolute value of total solar irradiance during the 2008 solar minimum period is $1360.8 \pm 0.5$ W m$^{-2}$ (Kopp and Lean 2011), and the change in total solar irradiance, $\Delta$TSI, of 1.3 W m$^{-2}$ (0.1%) in solar cycles 22 (1986–96) and cycle 23 (1996–2008) produced decadal solar radiative forcing $\Delta F_r = 0.7 \Delta$TSI/4 = 0.23 W m$^{-2}$ (Hansen and Lacis 1990). For comparison, radiative forcing by greenhouse gases over the 5-yr period of solar irradiance’s increase from solar minimum to maximum is less than 0.25 W m$^{-2}$. Such forcing estimates assume full adjustment of terrestrial climate processes to a new “equilibrium” state.
Absolute solar spectral irradiance is less well specified observationally, especially at infrared wavelengths (Meftah et al. 2018), than is its integral, the total solar irradiance. Spectral irradiance variability, which also depends (differently) on wavelength, is similarly less well specified observationally than is total solar irradiance variability. This is especially true over the solar cycle because drifts in spectral radiometer sensitivity during space operation can produce measurement uncertainties that are larger than the magnitude of the solar cycle changes (Lean and DeLand 2012; Mauceri et al. 2018). Conclusive detection of real spectral irradiance changes was achieved initially at ultraviolet wavelengths, where the relative changes are about an order of magnitude larger than at visible wavelengths. Spectroradiometers onboard the Solar Mesosphere Explorer (SME; 1980–89), UARS (1992–2005) (Lean et al. 1997; Rottman 2006), and SORCE (2003–present), as well as Solar Backscatter Ultraviolet (SBUV) instruments on multiple NASA and NOAA spacecraft (DeLand and Cebula 1998) all record spectral irradiance variability at wavelengths less than 400 nm. The launch of SORCE in 2003 successfully extended the detection of solar spectral irradiance variability to visible and near-infrared wavelengths (Rottman et al. 2005; Mauceri et al. 2018).

Figure 14-22 shows monthly mean variations in solar spectral irradiance in four broad wavelengths bands measured by SORCE since 2003 (green lines); 11-yr cycle variability is in the range 0.7% to 1.4% in the wavelength band 200–300 nm, 0.1% to 0.2% in the band 300–600 nm, 0.06% to 0.1% in the band 600–900 nm, and <0.05% in the band 900–1200 nm. Composite records of solar UV irradiance have been constructed by combining observations over multiple solar cycles (DeLand and Cebula 2008; Haberreiter et al. 2017), but their repeatability is generally insufficient to reliably specify the magnitude of solar cycle variability because of instrument sensitivity drifts and the lack of overlap needed to cross calibrate spectroradiometers with different absolute scales. The exception is the H I Lyman α emission line whose variability has been constructed since 1947 (Woods et al. 2000); solar Lyman α irradiance (at 121.5 + 0.5 nm) increased 3.5 mW m⁻² (60%) in solar cycle 22 and 2.5 mW m⁻² (40%) in cycle 24 (Snow et al. 2018).

c. Modeling solar irradiance variability

Climatological time scales, on which radiative forcing is typically defined, are considerably longer than the four decades of space-based solar irradiance observations. Models that relate the observed irradiance variations to historical indices of solar activity are therefore necessary to reconstruct solar radiative forcing prior to 1978. The primary causes of solar irradiance variability are dark sunspots and bright faculae, which, respectively, reduce and enhance the sun’s local radiative output (Foukal 1981), by different amounts at different wavelengths (Unruh et al. 2000; Lean et al. 2005). Solar rotation imposes a 27-day cycle on solar irradiance by altering the heliographic locations of dark sunspots and bright faculae on the disk, and the growth, transport, and decay of sunspots and faculae in response to a subsurface solar dynamo generates 11-yr irradiance cycles. Models that utilize indices of sunspot darkening and facular brightening reproduce the observed variations in total solar irradiance with high fidelity, including decreases up to a few tenths of a percent during solar rotation and cycle increases of 0.1% (Foukal and Lean 1988, 1990; Fröhlich and Lean 2004; Lean 2017).

Examples of solar irradiance variability models are those of the Naval Research Laboratory (NRLSSI2, NRLSSI2; Lean et al. 2005), which the NOAA CDR utilizes to estimate present and historical irradiance variations (Coddington et al. 2016), and the Spectral and Total Irradiance Reconstructions (SATIRE; Krivova and Solanki 2008; Krivova et al. 2010). The Naval Research Laboratory (NRL) models input a sunspot-darkening function calculated from direct observations of sunspot areas and locations on the sun’s surface and the Mg irradiance index facular proxy; multiple regression against observations determines the relative contributions of the two influences bolometrically and at individual wavelengths. The SATIRE model derives its two sunspot (dark sunspot umbra and penumbra) and two facular (bright faculae and network) inputs from solar magnetograms; a theoretical stellar atmosphere model specifies their wavelength-dependent contrasts relative to the background “quiet” sun (Unruh et al. 2000).

To assess the fidelity of such models by comparisons with extant, albeit imperfect, observations, Fig. 14-22 also shows monthly values of total solar irradiance and spectral irradiance in four broad wavelength bands according to the NRLSSI2 and SATIRE models. Figure 14-23 compares their total solar irradiance reconstructions over multiple solar cycles and corresponding spectral irradiance changes in selected epochs. Compared with the SATIRE model, the NRL model has a negligible downward trend during recent cycle minima (e.g., from 1986 to 2008 in Fig. 14-22), somewhat smaller solar cycle increases at near-ultraviolet wavelengths, and larger increases at longer wavelengths. In solar cycle 23, for example (Fig. 14-23), irradiance from 300 to 400 nm increases 0.4 W m⁻² in the NRLSSI2 model and 0.59 W m⁻² in the SATIRE model, while visible irradiance from 500 to 750 nm increases...
0.42 W m\(^{-2}\) in the NRLSSI2 model and 0.35 W m\(^{-2}\) in the SATIRE model.

Reliable historical reconstructions of solar irradiance depend on the availability of suitable sunspot and facular indices and on understanding plausible irradiance variability mechanisms. The lack of sunspots on the sun’s disk for several years during the Maunder Minimum (1645–1715) indicates anomalously low solar activity relative to the contemporary epoch (Eddy 1976). The possibility that solar irradiance was reduced during such periods relative to contemporary minima derives from the overall higher levels of \(^{14}\)C in tree rings and \(^{10}\)Be in ice cores (respectively) during the Spörer, Maunder, and Dalton minima. Cosmogenic isotope levels increase when solar activity decreases because the reduced solar magnetic flux in the heliosphere facilitates a great flux of galactic cosmic rays at Earth (McCracken et al. 2013). That cycles near 80, 210, and 2400 years manifest in cosmogenic isotope records of solar activity suggests the likelihood of similar periodicities in irradiance (e.g., Damon and Jirikowic 1992).

Initial estimates of the reduction in solar irradiance during the Maunder Minimum below contemporary solar minima considered two different scenarios (Lean et al. 1992; White et al. 1992). In one scenario, total irradiance decreased 1.5 W m\(^{-2}\) (about 0.1\%) due to the disappearance of faculae; a second scenario estimated a larger decrease of 2.6 W m\(^{-2}\) (about 0.2\%) because of an additional reduction in the background “quiet” sun, inferred from the reduced emission in noncycling sunlike stars (assumed to be in states of suppressed activity similar to the Maunder Minimum) relative to overall

---

**FIG. 14-23.** Reconstructed historical and projected future variations in total solar irradiance are shown (a) from 850 to 2300 according to two different models, the NRLTSI2 model (black lines) and the SATIRE model (orange lines, recommended for use in PMIP4). (b) The spectral irradiance changes for three selected periods, specifically solar cycle 23 (1996 to 2009), the Maunder Minimum (1645–1715) to modern maximum (1950–2009), and the Maunder Minimum to the medieval maximum (1100–1250).
higher emission in cycling stars. Questions about the applicability of sunlike stars for solar variability made these initial estimates speculative (Foukal et al. 2004). Current estimates of the solar irradiance increase from the Maunder Minimum to the present derive from a model of the transport of magnetic flux on the sun’s surface. The simulations suggest an increase of 0.5 W m$^{-2}$ (about 0.04%) in total solar irradiance at cycle minima over the past ~300 years, from the accumulation of magnetic flux during successive 11-yr cycles of increasing strength (Wang et al. 2005). This estimate of long-term solar irradiance variability is a factor of 5 smaller than inferred from sunlike stars [Table IV in Lean et al. (2005) summarizes estimates of total solar irradiance reduction in the Maunder Minimum]. Figure 14-23 shows reconstructions of total solar irradiance from 850 to 2300 using the $^{13}$C cosmogenic isotope record of Roth and Joos (2013), assuming a reduction in the Maunder Minimum of 0.5 W m$^{-2}$. Figure 14-23 also shows spectral irradiance changes in the NRLSSI2 and SATIRE models from the Maunder Minimum to the present and to the medieval maximum.

Both the magnitude and temporal structure of longer-term irradiance changes remain uncertain. The only direct index of solar activity prior to 1882 is the sunspot number, which is undergoing renewed scrutiny and debate (Clette and Lefèvre 2016; Kopp et al. 2016). The two historical irradiance reconstructions in Fig. 14-23 differ notably prior to 1882 because of their different parameterizations of irradiance in terms of sunspot numbers and cosmogenic isotopes (Lean 2018a). The relationship between solar irradiance and cosmogenic isotopes is complex and poorly known, in part because the magnetic fields that produce sunspots and faculae at the sun’s surface are not the same as those that modulate galactic cosmic rays in the heliosphere (Lean et al. 2002). As well, distinctly different terrestrial processes produce cosmogenic archives in tree rings and ice cores (Delaygue and Bard 2011; Steinhilber et al. 2012; Roth and Joos 2013).

d. Climate response to solar radiative forcing

Just as the detection of terrestrial responses to solar activity initially signified the relevance of solar radiative forcing for understanding climate change, in lieu of direct observations of the forcing itself, so too does ongoing analyses of ever-lengthening terrestrial observations and newly extracted, high-fidelity paleoclimate records continue to strengthen and expand the evidence. Using indicators such as sunspots and cosmogenic isotopes to identify times of high and low solar activity during the 11-, 80-, and 210-yr cycles, solar-related changes are identified in diverse climate parameters that range from low-latitude drought and rainfall (e.g., Verschuren et al. 2000; Neff et al. 2001; Haug et al. 2003; Antico and Torres 2015), associated with intertropical convergence zone displacement (Novello et al. 2016) and a La Niña-type response in the tropical Pacific (Mann et al. 2005), to mid- and high-latitude “centers of action” (Christoforou and Hameed 1997), storm tracks, and winter intensity (e.g., Barriopedro et al. 2008; Mann et al. 2009; Lockwood et al. 2010), associated with the North Atlantic Oscillation and the circumpolar vortex.

On global scales, climate signals related to the 11-yr solar cycle were detected first in basinwide ocean temperatures (White et al. 1997) then in global lower tropospheric temperature (Michaels and Knappenberger 2000). Observational temperature and ozone databases are now sufficiently long that statistical analyses readily isolate in them solar responses, both globally and regionally, from other concurrent influences (Douglass and Clader 2002; Lean and Rind 2008; Foster and Rahmstorf 2011). Figure 14-24 shows the solar cycle component, thus extracted, of <0.1°C in global surface temperature and <3 DU (1%) in total ozone compared with natural [volcanic, El Niño–Southern Oscillation (ENSO), quasi-biennial oscillation (QBO)] and anthropogenic (greenhouse gases and ozone-depleting substances) components, and Fig. 14-25 shows the corresponding geographical response patterns. These estimates of climate’s response to solar forcing are attained by linearly regressing indices of the simultaneous natural and anthropogenic influences against ( deseasonalized) monthly mean surface temperature (Fig. 14-24a) and total ozone (Fig. 14-24b) observations from 1979 to 2017 (Lean 2017, 2018b).

Time-dependent simulations of climate’s response to solar radiative forcing on climatological time scales became possible with the reconstruction of historical solar irradiance. Simulations using energy balance models initially suggested that the global surface temperature response to reconstructed solar irradiance cycles since 1874 (Foukal and Lean 1990) was likely undetectable, the transient response of 0.03°C being notably smaller than the equilibrium response because of attenuation (~80%) by the thermal inertia of the ocean (Wigley and Raper 1990). But subsequent analysis of historical surface temperature observations did detect a solar cycle response of 0.06°C by statistically extracting the modeled spatial pattern of the response to the forcing (Stevens and North 1996).

The first general circulation model simulations of climate’s response to time-dependent solar radiative forcing found a global surface temperature increase of ~0.5°C since the Maunder Minimum (Cubasch et al. 1997; Rind et al. 1999). Decreased solar irradiance
during the Spörer, Maunder, and Dalton solar activity minima (Eddy 1976) and enhanced volcanic activity are posited causes of anomalously cold surface temperatures from ~1300 to 1850, during the Little Ice Age (e.g., Mann et al. 2005, 2009). Even though the simulations input a factor of 5—or more—larger increase in total solar irradiance (Lean et al. 1995) than current estimates (Wang et al. 2005), they nevertheless identified that water vapor feedbacks, cloud-cover changes, and land–sea contrasts contribute to the surface response to solar radiative forcing, with enhanced warming in subtropical regions similar to that forced by increasing greenhouse gas concentrations. The simulations further established that variations in solar irradiance were unlikely to be the primary cause of global warming in the postindustrial period, as some statistical correlations between solar cycle length and Northern Hemisphere temperature had suggested (Friis-Christensen and Lassen 1991).

State-of-the-art general circulation models now include couplings between the land, ocean, and atmosphere, functional middle atmospheres with ozone chemistry, and the ability to input realistic solar spectral irradiance changes. Analyses of ensembles of simulations made with various such models, designed to isolate responses of different terrestrial regimes to solar radiative forcing, demonstrate both a direct response of the

---

**Fig. 14-24.** Shown as the black lines are observed changes from 1979 to 2017 in (a) monthly averaged global surface temperature and (b) monthly averaged global total ozone. Also shown by the blue lines in (a) and (b) are the changes in the respective observations according to statistical models. The statistical models are constructed by using linear least squares regression of the observations against indices of the known sources of their variability. The relative contributions of the individual components to the observed changes are identified in the lower two panels; these include ENSO, QBO, volcanic aerosols, the solar irradiance cycle, changes in the concentrations of anthropogenic GHG, and the EESC of ozone-depleting substances (adapted from Lean 2017, 2018b).
land and ocean, dependent in part on the regional distribution of clouds, and an indirect response facilitated by stratospheric ozone and temperature changes (Rind et al. 2008; Meehl et al. 2009a). Convective and dynamical processes disperse the forcing geographically and altitudinally, altering extant dynamical patterns such as the Hadley and Walker circulations and impacting, in particular, the hydrological cycle. IPCC’s AR5 climate assessment included simulations made with 13 models that resolve the stratosphere (Mitchell et al. 2015), 6 of which include interactive ozone chemistry (Hood et al. 2015). Modeled responses to solar cycle irradiance changes are evident at the surface, in the ocean (Misios et al. 2015), and in the troposphere, stratosphere, and ozone layer (Hood et al. 2015). While the simulated responses are generally of smaller magnitude than in observations (e.g., global-mean surface warming of 0.07°C), the processes and patterns are qualitatively similar, including changes in precipitation and water vapor leading to weaker Walker circulation (Misios et al. 2015) and a stratosphere-related North Atlantic surface response (Mitchell et al. 2015).

Figure 14-25 compares statistically extracted geographical patterns of the terrestrial response to solar radiative forcing with estimates made by a physical climate model (Rind et al. 2008). Differences between the physical and statistical model patterns suggest that deficiencies remain in one or both. Uncertainties in the hundreds of parameterizations that seek to account for the multiple integrated processes that heat the land and

---

**Fig. 14-25.** (top) The annually averaged relative distribution of received solar radiation at Earth. The regional patterns of terrestrial responses to changes in solar radiation during the 11-yr cycle, statistically extracted from observations (Lean 2017, 2018b), are shown (a) for total ozone and (b) for surface temperature. (c),(d) For comparison, the terrestrial responses to the solar cycle simulated by a general circulation model (GISS Model 3; Rind et al. 2008) are also shown.
ocean, and redistribute this heat regionally and vertically, compromise physical model simulations. Statistical models suffer from uncertainties in the predictors and covariance among them (such as between solar and anthropogenic indices), including distinguishing whether covariance is physically based or random. The limited duration of the most reliable observations and indices exacerbate such uncertainties. Articulating and reconciling differences between the statistical and physical models is expected to improve understanding of processes that facilitate terrestrial responses to solar radiative forcing and may help improve physical model parameterizations of these processes. It is increasingly apparent that solar radiative forcing initiates a continuous spectrum of coupled interactions throughout Earth’s land, ocean, and atmosphere on multiple time scales with different and interrelated regional dependencies. Differential heating of the land and oceans, equator and poles, and surface and atmosphere drive these responses; the processes involved are those by which climate responds to other radiative forcings, including by increasing greenhouse gas concentrations, albeit with different magnitudes during the 11-yr solar cycle. Global terrestrial responses to solar radiative forcing have detectable magnitudes during the 11-yr solar cycle. Global physical processes that connect variations in solar irradiance and irradiance variations require resolution and dynamical responses to this forcing, and the modulation of extant circulation patterns throughout the integrated system.

2) UNCERTAINTIES

Not yet known with the needed certainty is the magnitude of spectral irradiance changes in the solar cycle. This is because observations lack the long-term stability to establish this unequivocally, and models disagree about the apportioning of the changes to near-UV versus visible–near-IR wavelengths. Less certain still are the magnitudes of multidecadal irradiance variations and their possible mechanisms. Yet to be proven is the assumption in historical reconstructions that solar irradiance varies on time scales longer than the 11-yr activity cycle and whether observations of sunlike stars can provide useful estimates of the magnitude of this variability.

Physical processes that connect variations in solar irradiance and cosmogenic isotopes, including modulation by solar magnetic flux, the flow of galactic cosmic rays through the heliosphere, and production of isotopes in terrestrial archives are conceptually established but not yet quantified with the needed certainty. Similarly, the terrestrial processes and model parameterizations thereof that facilitate the multiple pathways that transform solar radiative forcing to climate variability are generally recognized, but their specifications require validation and improvement. This includes the deposition of incident solar spectral energy, direct chemical and dynamical responses to this forcing, and the modulation of extant circulation patterns throughout the integrated system.

3) CHALLENGES

The highest priority going forward is the continuous monitoring of solar irradiance with the highest possible accuracy and repeatability to extend the extant record of solar radiative forcing, exemplified by the launch in 2018 of the TSIS on the International Space Station.

Differences among observed and modeled absolute irradiance and irradiance variations require resolution, including the magnitude of interminima changes in the space era and the spectral dependence of the variability.

Physical climate models of the future are challenged to fully capture and parameterize the multiple pathways by which solar radiation enters and alters the integrated terrestrial environment, including under different conditions of other natural and anthropogenic forcings.

The reconciliation of the magnitude, pattern, and time lags of terrestrial responses to solar radiative forcing extracted statistically from observations with those calculated by physical models may aid the pursuit of this challenge.

e. Summary: Successes, uncertainties, challenges

1) SUCCESSES

Both measurements and models have now established that solar irradiance varies at all wavelengths, with different magnitudes at different wavelengths. Total (spectrally integrated) solar irradiance increased $1.3 \pm 0.2\,\text{W m}^{-2}$ (0.1%) in solar cycle 23 (1996–2008), producing radiative forcing of 0.22 W m$^{-2}$. Spectrally, the energy change maximizes at 300–400 nm, which increased 0.4 W m$^{-2}$ in solar cycle 23.

Solar irradiance variability is a result of the sun’s magnetic activity, which alters radiative output locally in dark sunspots and bright faculae. Models of the net, global influence of sunspots and facular reproduce observed total solar irradiance variability with high fidelity on time scales of the sun’s 27-day rotation.

Observations and models of Earth’s surface and atmospheric temperature and ozone amount indicate that terrestrial responses to solar radiative forcing have detectable magnitudes during the 11-yr solar cycle. Global responses are $\sim 0.1^\circ\text{C}$ in surface temperature, $\sim 0.3^\circ\text{C}$ in lower-stratospheric temperature, and $\sim 3\,\text{DU}$ (1%) in total ozone; the response patterns are regionally inhomogeneous and differ from that of the incident solar radiative forcing. There is abundant terrestrial evidence in paleoclimate records and solar activity proxies for solar radiative forcing with cycles near 80 and 210 years, in addition to the 11-yr cycle.
9. Stratospheric aerosols

a. Introduction

An important discovery of the twentieth century is that some amount of submicron sulfate particles are permanently present in the stratosphere perturbing Earth’s radiative balance, climate, and weather (Junge et al. 1961; Turco et al. 1982; Pueschel 1996; Hamill et al. 1997). The abundance of stratospheric aerosols greatly increases after explosive volcanic eruptions that inject materials directly in the stratosphere. The tremendous success in observations and theoretical understanding of stratospheric aerosols achieved during recent decades is briefly reviewed in this section.

Volcanic hazards have been documented since antiquity. Pompeii in the ancient Roman Empire was destroyed by the eruption of Vesuvius in AD 79 (Zeilinga de Boer and Sanders 2002). Plutarch mentioned that the Etna eruption in 44 BC dimmed the sun and killed the crops, causing famine in Rome and Egypt (Forsyth 1988). The systematic compilation of active volcanoes and past volcanic eruptions started in the mid-nineteenth century (Scope 1862). Coming to the present times, there is now available a comprehensive database of active volcanoes (Simkin et al. 1981; Simkin 1993).

It was long suspected that explosive volcanic eruptions affect the weather, climate, and human health through the injection into the atmosphere of large amounts of solid particles (i.e., volcanic ash) and gases (Coakley 1981; Robock 2000; Timmreck 2012; Stenchikov 2016). Benjamin Franklin, then a U.S. ambassador to the court of Louis XVI, related the 1783 Laki eruption in Iceland to a pioneering balloonist Leon Tisere de Bort in 1902 (Greene 2000). Currently, the stratospheric aerosol layer is often referred to as Junge layer both for volcanically quiescent and active periods, although the aerosol abundance, vertical extent, and horizontal spread drastically change after an explosive volcanic eruption.

In the volcanically quiescent periods, the background Junge-layer height varies around 20 km and is modulated by the QBO. The aerosol abundance is maintained by the mostly tropical cross-tropopause transport (Fueglistaler et al. 2009) of sulfur-containing gases, carbonyl sulfide (OCS) (Crutzen 1976), and SO2 (Brühl et al. 2012; Sheng et al. 2015), as well as aerosols of natural and anthropogenic origin (Brock et al. 1995). Deep convective cells could also overshoot tropospheric materials into the stratosphere (Stenchikov et al. 1996). The much-debated contribution of anthropogenic sulfur from highly polluted East Asia is found to be of minor importance (Deshler et al. 2006; Vernier et al. 2011; Thomason and Peter 2006). This is consistent with the observation that despite the increase in anthropogenic emissions there is no measurable long-term trend in background aerosols (Deshler et al. 2006; Vernier et al. 2011; Thomason and Peter 2006) during 1970–2005. The estimated total net flux of sulfur into the stratosphere is about 180 GgS yr⁻¹, but this figure has an uncertainty of at least 50% (Thomason and Peter 2006; Kremser et al. 2016). The input of other aerosols, meteoric from the middle atmosphere, organic from troposphere, and mixed from air traffic and rocket exhaust, is little by mass but may affect aerosol microphysics (Turco et al. 1982; Bardeen et al. 2008; Kremser et al. 2016; Gomez Martin et al. 2017).

The diameters of background aerosol particles range from 10 to 100 nm, and their total optical depth is on the order of 0.001 in visible (Vernier et al. 2011), resulting in radiative forcing (compared to zero concentrations) of about −0.01 to −0.05 W m⁻². So, although background stratospheric aerosols are important indicators of the stratosphere–troposphere chemical exchange and chemical processes in the stratosphere itself, their radiative effect is relatively small but not completely negligible. Note that radiative forcing in the present paper is with reference to preindustrial times when there was likely a small but nonzero background concentration.

However, the pure background Junge layer is rarely observed as it is frequently perturbed by explosive volcanic eruptions that directly inject into the lower-stratosphere volcanic ash, sulfur-containing gases, mostly SO2 and H2S, water vapor, CO2, halogens, nitrogen (N2), and other species. After such emissions, the thermal and chemical relaxation of the stratosphere to background level takes 7–8 years (Brasseur and Granier 1992; Thomason and Peter 2006). So during the observation period that started in the 1970s, there are only a
few time windows when the background Junge layer could be sampled.

Volcanic ash particles, mainly recondensed silicates, usually exceed 2 μm in diameter. Although ash immediately after eruption develops a measurable radiative forcing, it does not produce a long-term climate effect as it gravitationally deposits in about a week or two. The finest ash particles might be present in the stratosphere for a few months, but their radiative effect is negligible (Niemeier et al. 2009; Guo et al. 2004a,b).

The model studies suggest that the aerosol plume from a strong equatorial volcanic eruption could heat the tropical tropopause layer (TTL) facilitating the tropospheric water vapor penetration into the stratosphere (Joshi and Shine 2003; Robock et al. 2009; Löfller et al. 2016). The heating of the TTL after volcanic eruptions appears challenging to detect in observations even for the most recent strong eruption of Mount Pinatubo (Fueglistaler et al. 2013; Randel et al. 2004; Chiu et al. 2006). This is because the TTL temperature and water vapor flux into the stratosphere are also affected by QBO, ENSO, and the strength of the Brewer–Dobson circulation. Dessler et al. (2014) used multiple regression analysis to account for all those factors and have shown the increase of the water vapor mixing ratio in the air entering the tropical stratosphere both for El Chichon and Pinatubo eruptions.

Volcanic SO2 and H2S are oxidized in the stratosphere by a photochemically produced hydroxyl radical to form sulfate aerosols with a characteristic conversion time of about one month (Bluth et al. 1992, 1993; Read et al. 1993), although this rate could vary at different stages of the process (LeGrande et al. 2016). Initially, aerosol particles are formed due to binary nucleation of sulfuric acid and water vapor and are subject to coagulation and diffusional growth as well as gravitational settling (Turco et al. 1982; Pueschel 1996; Hamill et al. 1997). They exert substantial perturbation of the radiative energy budget of the planet (Lambert et al. 1993; Baran and Foot 1994; Minnis et al. 1993; Barnes and Hoffman 1997; Lacic et al. 1992; Stenchikov et al. 1998).

Volcanic aerosols are dispersed in the stratosphere by wave-driven Brewer–Dobson circulation (Holton et al. 1995) modulated by the QBO phase (Trepte and Hitchman 1992) to be deposited at high latitudes. The stratospheric e-folding residence time of equatorial injections with respect to Brewer–Dobson transport is about two years (Hamill et al. 1997). The lifetime of high-latitude volcanic injections is somewhat shorter than the low-latitude ones because of proximity to a pole and absence of the slowly emptying aerosol equatorial reservoir blocked by the subtropical barrier (Oman et al. 2006b). For large eruptions, gravitational settling of quickly growing sulfate particles intensifies aerosol removal restricting the magnitude of the climate impact of supereruptions (Pinto et al. 1989; Timmreck et al. 2010). Volcanic aerosols deposited to Antarctica and/or Greenland snow affect the chemical composition of ice, thus recording the history of Earth’s volcanism for thousands of years (Zielinski 2000; Cole-Dai 2010).

c. Observations of stratospheric aerosols

The first generation of aerosol-viewing instruments provided invaluable empirical knowledge about stratospheric loadings and optical properties, however, with some significant gaps in space and time (Dessler 2008; Deshler et al. 2003; Hofmann et al. 1975; Baumgardner et al. 1992; Borrman et al. 2000; Fiocco and Grams 1964; Stothers 1996, 1997, 2001a,b; Krueger et al. 2000; Carn et al. 2003; McCormick 1987; Antuña et al. 2003; Thomason and Taha 2003; Randall et al. 2000, 2001). Thomason and Peter (2006) and Kremser et al. (2016) overview extensively the available observations.

The recognized deficiency of existing observations is that aerosol size distribution, which affects both aerosol optical characteristics and sedimentation velocities, suffers from significant retrieval uncertainties (Kremser et al. 2016; Bingen et al. 2004a,b; Bourassa et al. 2008; Malinina et al. 2018). During the recent decade, a new generation of instruments for monitoring aerosols and precursor gases has emerged. Among them are the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), Ozone Mapping Profiler Suite (OMPS), Ozone Monitoring Instrument (OMI), Infrared Atmospheric Sounding Interferometer (IASI), CALIPSO, Cloud–Aerosol Transport System (CATS), Optical Spectrograph and Infrared Imager System (OSIRIS), and Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY). OMI and OMPS continue the Total Ozone Mapping Spectrometer (TOMS) measurement record providing SO2 loadings used to document the global volcanic degassing (Carn et al. 2016). MIPAS sees the vertically resolved SO2 and aerosol volume (Höpner et al. 2013, 2015). The vertically resolved aerosol extinctions are detected by the limb-profiling SCIAMACHY (Burrows et al. 1995; Bovensmann et al. 1999; von Savigny et al. 2015), OSIRIS (Bourassa et al. 2007), and lidar instruments, CATS (Yorks et al. 2015) and CALIOP (Vernier et al. 2012). The new limb-scattering instruments observe aerosol plume more frequently than those based on the solar occultation technique and make it possible to reliably retrieve more parameters of aerosol particle size distribution (PSD) than were possible in the past (Malinina et al. 2018). For the forcing calculations, it is important to smoothly merge the past and current aerosol observations.
to produce seamless datasets of stratospheric aerosol parameters for an extended period of time (Thomason et al. 2018).

d. Radiative forcing of stratospheric aerosols

Stratospheric aerosols, both volcanic and background, scatter the incoming shortwave radiation depleting the direct and enhancing the diffuse downward solar fluxes; they also absorb shortwave near-infrared and absorb and emit outgoing terrestrial radiation. The cumulative radiative effect of stratospheric aerosols is to cool Earth’s surface and heat the aerosol layer in the lower stratosphere.

Volcanic eruptions that have historically exerted the strongest radiative forcing have (i) significant SO$_2$/H$_2$S injected into the stratosphere (although there is growing evidence of nonlinearity of injections strength and radiative forcing; e.g., Niemeier and Tilmes 2017), (ii) tend to occur in tropical regions where both hemispheres of the globe are impacted by the subsequent perturbation to the aerosol optical depth, and (iii) inject SO$_2$ to sufficiently high altitudes within the stratosphere (e.g., Jones et al. 2016a).

The large perturbations of Earth’s radiative balance caused by explosive volcanic eruptions (e.g., Pinatubo) are discernible in observations; however, this does not lend itself readily to quantifying their actual radiative forcing (Dutton and Christy 1992; Minnis et al. 1993; Russell et al. 1993). The theoretical calculations of the radiative forcing of stratospheric aerosols were first attempted using conceptual models (Lacis et al. 1992; Harshvardhan 1979; Toon and Pollack 1976). Because aerosol microphysical and optical characteristics, which have to be compiled from observations or calculated within the model, are the major input into the radiative forcing calculations, we discuss both these aspects together here.

The first generation of the atmospheric general circulation models simulated the impact of volcanic aerosols using simplified approaches, that is, assuming a reduction of the solar constant, increase of planetary albedo, or representing stratospheric aerosols by a single reflecting layer (e.g., Broccoli et al. 2003; Soden et al. 2002).

The existing aerosol observations were used to build the global aerosol datasets with precalculated aerosol optical/microphysical characteristics that could be implemented in climate models (Stenchikov et al. 1998; Stenchikov 2016; Ramachandran et al. 2000; Sato et al. 1993; Hansen et al. 2002; Schmidt et al. 2011; Tett et al. 2002; Ammann et al. 2003). One approach is to use the observed/reconstructed aerosol optical depth (usually in visible) and assume aerosol composition and size distribution to calculate the aerosol extinction, single-scattering albedo, and asymmetry parameter required for radiative transfer models as input (Stenchikov et al. 1998; Sato et al. 1993). Another approach uses the empirical estimates of SO$_2$ emissions and a simplified model to distribute them globally and to obtain the aerosol optical parameters (Ammann et al. 2003; Gao et al. 2008). Ammann et al. (2003) and Sato et al. (1993) datasets have essentially provided the bases for implementing volcanic aerosols in virtually all of the climate models that have performed the twentieth century climate integrations within IPCC AR4 (Stenchikov et al. 2006; Forster et al. 2007).

For the IPCC AR5 and CMIP6, the improved “gap filled” SAGE II, version 6, aerosol product from Thomason and Peter (2006) was employed (Arfeuille et al. 2013; Zanchettin et al. 2016). All three stratospheric optical depths (SATO, Ammann, and CMIP6) in Fig. 14-26 vary by about 30%, with Amman’s optical depth being the largest and CMIP6 being the smallest.

![Fig. 14-26. Global-mean optical depth of stratospheric sulfate aerosols for 0.55 μm calculated using CMIP6, Sato et al. (1993) with (Schmidt et al. 2011) corrections, and Ammann et al. (2003) datasets.](image-url)
Figures 14-27 and 14-28 compare the all-sky SW, LW, and SW + LW instantaneous radiative forcing at the top of the atmosphere and perturbations of heating rates calculated using SATO and CMIP6 inputs within the GFDL CM2.1 (Delworth et al. 2006) employing a double radiation call. To calculate optical characteristics of stratospheric aerosols for the SATO case, it was assumed that the aerosol has lognormal distribution with the time- and latitude-varying effective radii and a fixed geometric width of 1.8 $\mu$m (SATO1.8) or 2.0 $\mu$m (SATO2). Despite the differences in the input information and assumptions, the changes in total radiative balance for the three datasets appear to be quite close. Both SATO’s datasets slightly overestimate the SW radiative forcing in comparison with Minnis et al. (1993). The CMIP6 heating rates appear to be higher than expected (Stenchikov et al. 1998) and shifted toward the SW heating. Typical stratospheric sulfate particles absorb SW radiation only in near-IR starting from 2.5 $\mu$m, where solar flux is weak. This is why LW heating is expected to prevail, contributing about 70% of the effect (Stenchikov et al. 1998). The stratospheric heating is important, as it controls stratospheric dynamic responses (Ramaswamy et al. 2006a).

The complexity of radiative, microphysical, and transport processes forced by volcanic aerosols suggests that
it is important to calculate aerosol radiative effects interactively with the aerosol plume development rather than use a precalculated set of aerosol optical parameters. To accomplish this, it is necessary to know the SO2 volcanic emissions (Krueger et al. 2000; Hõpfer et al. 2013, 2015) and be able to calculate development, transport, and decay of a volcanic aerosol layer.

The “bulk” aerosol models calculate SO2 to H2SO4 conversion and transport their bulk concentrations. Sulfate aerosols are assumed to form instantaneously with the prescribed size distribution (Timmreck et al. 1999; Oman et al. 2006a; Aquila et al. 2012) that defines aerosol optical properties and deposition rates. Modal aerosol models keep track of aerosol number density approximating the aerosol size distribution by a few lognormal modes with the prescribed width and varying modal radii, accounting for coagulation, condensation growth, and size-dependent gravitational settling (Niemeier et al. 2009; Brühl et al. 2015; Dhomse et al. 2014; LeGrande et al. 2016; Sekiya et al. 2016). The aerosol sectional microphysical models are the most accurate but computationally more demanding (English et al. 2013; Mills et al. 2016).

There are still significant discrepancies between models and between the models and observations. This remains a challenging issue. The 1991 Pinatubo case study is an important test bed where different approaches have been compared and could be further investigated. For example, in Brühl et al. (2015), the aerosol optical depth relaxes too fast, but in Mills et al. (2016), the stratospheric aerosol plume decays too slowly, and the initial SO2 loading has to be decreased by almost a factor of 2 to make the results consistent with observations.

**e. Small volcanoes, climate hiatus, and geoengineering analogs**

The slowing of global warming, or climate hiatus, in 2000–13, despite continued emission of greenhouse gases, attracted widespread attention (Meehl et al. 2011; Myhre et al. 2013). Many mechanisms were suggested as causal factors. These included natural variability associated with increased ocean heat uptake (Balmaseda et al. 2013) and cooling forced by small volcanic eruptions (Fyfe et al. 2013; Haywood et al. 2014; Santer et al. 2014; Solomon et al. 2011). The latter refers to eruptions of Kasatochi in August of 2008, Sarychev in June 2009, and Nabro in June 2011. They were the most significant recent events, however, 15–20 times weaker in terms of SO2 injection than for the Pinatubo eruption. The estimated global-mean surface temperature perturbations that they could cause range from 0.02 to 0.07 K (Haywood et al. 2014; Santer et al. 2014). However, Andersson et al. (2014) reported that about 30% of aerosols from these small volcanoes were retained in the lowermost stratosphere and their total optical depth was underestimated in observations.

The deliberate injection of aerosols and aerosol precursors in the lower stratosphere suggested to reduce greenhouse warming (Crutzen 2006; Wigley 2006; Govindasamy and Caldeira 2000; Robock et al. 2010; Heckendorn et al. 2009) is discussed in detail in section 13. The associated processes have much in common with the effects of volcanic aerosols. Therefore, the understanding of all aspects of stratospheric aerosols and climate links, which we gain from investigating the climate consequences of volcanic eruptions, is important for a feasibility assessment of solar management schemes (Pitari et al. 2014; Tilmes et al. 2009; Aquila et al. 2014; Haywood et al. 2013; Aquila et al. 2012; Kravitz et al. 2013a; Tilmes et al. 2013).

**f. Dynamic and thermal responses to volcanic eruptions**

Improvements in our understanding of volcanic forcing help to better understand past climate and make a better climate prediction. It also enables the radiative forcing and accompanying transient response due to volcanic aerosols to be placed in perspective, relative to the forcing and responses due to the increases in the anthropogenic well-mixed greenhouse gas emissions. Since 1850, volcanic forcing has offset the ocean heat content increase due to the global-mean warming by about 30% (Delworth et al. 2005). Comparison of simulated and observed climate responses to the major volcanic eruptions helps to evaluate volcanic forcing itself. The relatively large transient forcing by volcanic aerosols offers a platform to test climate model simulations of stratospheric and surface temperature perturbations against observations.

The net radiative effects of volcanic aerosols on the thermal and hydrologic balance (e.g., surface temperature and moisture) have been highlighted in (Kirchner et al. 1999; Free and Angell 2002; Jones and Mann 2004; Trenberth and Dai 2007). Atmospheric temperature after volcanic eruptions relaxes for 7–10 years, while the deep ocean retains a thermal perturbation for about a century (Stenchikov et al. 2009; Delworth et al. 2005). Gregory et al. (2013) indicated the importance of the preindustrial volcanic forcing to predict future climate correctly. The prolonged volcanic activity could be a reason for a long-term climate cooling as it had arguably happened during the medieval Little Ice Age in 1300–1850 (Free and Robock 1999) when in the middle of this period the cooling was enhanced by the Maunder Minimum in solar irradiance (Eddy 1976).

In addition, the differential heating/cooling due to volcanic aerosols affect atmospheric circulation. It is
believed these circulation responses could cause a positive phase of the Arctic oscillation and winter warming in high northern latitudes (Ramaswamy et al. 2006a; Shindell et al. 2003, 2004; Stenchikov et al. 2002, 2004, 2006; Perlwitz and Graf 2001; Toohey et al. 2014), prolonged or even initiate El Niño (Adams et al. 2003; Pausata et al. 2015; Predybaylo et al. 2017; McGregor and Timmermann 2011; Ohba et al. 2013), or damp monsoon circulations (Trenberth and Dai 2007; Anchukaitis et al. 2010; Iles et al. 2013; Schneider et al. 2009). There are still large discrepancies between the models on the magnitude and the leading mechanism that forces those dynamic responses, and observations are not long enough to provide empirical proof of a concept. For example, Polvani et al. (2019) argued that the positive phase of the Arctic Oscillation in the winter of 1991/92 was not casually forced by the 1991 Pinatubo eruption, as it was not associated with the strong northern polar vortex. However, one has to take precaution making a far-reaching conclusion from their analysis as the authors only considered one volcanic winter that does not exhibit a statistically significant climate signal.

One robust finding in terms of dynamical response to high-latitude eruptions that preferentially load one hemisphere relative to the other is that tropical precipitation associated with the intertropical convergence zone is shifted toward the unperturbed hemisphere in both observations and global climate models (Oman et al. 2005; Haywood et al. 2013). Thus, significant high-latitude eruptions in the Northern Hemisphere (e.g., Katmai, which erupted in 1913) can lead to drought in sub-Saharan Africa and cause the North Atlantic hurricane frequency to dramatically reduce in years subsequent to the eruptions (Evan 2012; A. C. Jones et al. 2017). These impacts are relatively well understood from theoretical constraints on cross-equatorial energy and moisture transport (e.g., Bischoff and Schneider 2014, 2016). Equatorial eruptions also can affect the position of African rain belt by the combined effect of the preferential hemispheric summer cooling and damping of Indian monsoon (Dogar et al. 2017).

g. Summary

Stratospheric aerosols exert a substantial, albeit transient, impact on climate after the Junge layer is replenished by strong volcanic injections. For the equatorial eruptions, the radiative forcing peaks in about a half a year after a volcanic explosion and relaxes with the e-folding time of 1–2 years. For the high-latitude eruptions, the e-folding time is shorter than for tropical ones. Despite the transient nature of the volcanic forcing, the global ocean integrates the cooling from multiple eruptions, extending the climate response to decades and even centuries (Delworth et al. 2005; Stenchikov et al. 2009).

Our understanding of the effect of stratospheric aerosols has grown substantially over the last century, from descriptive and intuitive knowledge base to the full-scale first-principle modeling supported by ground-based and satellite observations. Despite this progress, the error bars in volcanic radiative forcing probably remain larger than 20%–30%. Because we have a limited ability to reconstruct volcanic forcing in the past, it is extremely important to further develop models that could interactively simulate volcanic plume development and its radiative effect. The best models so far demonstrate a sizable discrepancy with available observations that also may bear a significant uncertainty. One important bottleneck is aerosol particle size distribution that is controlled by finescale microphysical processes. Particle sizes are important as they define both radiative effects of aerosols and their lifetime with respect to gravitational settling. The accumulation of the effect of small volcanic eruptions has to be better understood as it contradicts the expectation of a smaller lifetime of above-tropopause emissions. The precalculated, based on observations, aerosol datasets have their value in helping to better calibrate simulated climate responses to volcanic forcing.

It is important to consider radiative forcing and climate responses in combination, as this gives important feedback on how well a model reproduces the observed climate variations. The climate models are capable of calculating the thermodynamic responses to the volcanic aerosols forcing but fail to consistently reproduce the circulation anomalies forced by volcanic eruptions. Further development of model capabilities and stratospheric aerosol monitoring are necessary to reduce uncertainties in past and future climate simulations.

10. Total natural and anthropogenic radiative forcing

This section describes developments in the comparison of different forcing agents, which then naturally leads to estimates of the total forcing and its time evolution, and thus acts as a synthesis of the material in the previous sections. The nature of the forcing agents is very different in terms of magnitude, uncertainty, spatial distribution, and time evolution (see section 2–9).

a. Complexity in comparing forcing agents

Kiehl and Briegleb (1993) were the first to provide model estimates of the geographical distribution of the net forcing of WMGHGs and the direct sulfate aerosol effect. The results showed that the WMGHG forcing in
northern industrialized regions at midlatitudes was strongly offset by the direct aerosol effect, but the offset was much weaker in the tropics where WMGHG forcing is at maximum. A distinct spatial distribution was also found in modeling studies for tropospheric ozone and biomass burning by Penner et al. (1992) with a maximum near the emission regions located over the continents and for stratospheric ozone forcing by Ramaswamy et al. (1992) with a negative forcing in the mid- to high latitudes but near zero or small negative in the low latitudes.

Geographical distributions of various climate drivers, and further estimates of RF and their uncertainties, were given in Shine and Forster (1999) and Hansen et al. (2000) and assessed in TAR. All these estimates showed a large difference in the uncertainty of WMGHG forcing and other climate drivers, in particular, aerosol effects. The much larger uncertainties are due to known factors such as the aerosol distribution and optical properties, as well as uncertainties in the physical process of aerosol–cloud interactions. Compared to relatively low uncertainties for WMGHGs, this made it difficult to provide a net RF. In addition to providing uncertainty ranges for the climate drivers, TAR presented the “Level of Scientific Understanding,” which showed large difference among the climate drivers. This illustrated the difficulty at that time of providing a net forcing for both global and annual means and geographical distributions. Ramanathan et al. (2001) and Kaufman et al. (2002) on the regional scale and AR4 globally pointed out the distinct differences in the net anthropogenic forcings at TOA and surface, which demonstrates the sharp differences between the relatively homogeneous WMGHG forcings and the much more spatially inhomogeneous aerosol forcing.

b. Applications of probability distribution functions to derive total anthropogenic forcing

Boucher and Haywood (2001) provided a method to estimate a net global-mean RF from components with different uncertainties, a method that has since been used in IPCC assessments beginning with AR4. The method used probability distribution functions (PDFs) for the individual climate drivers and a Monte Carlo approach to estimate the net RF. In Boucher and Haywood (2001), various assumptions on the shape of the PDFs such as normal and lognormal distributions were investigated. They found a higher sensitivity to how uncertainty ranges should be interpreted than to the shape of the PDFs. This has led to an improved quantification of uncertainty range and confidence levels in later research and whether RF numbers are given as, for example, one standard deviation or 5%–95% confidence intervals. The method of Boucher and Haywood (2001) allows the calculation of a mean net RF (with an uncertainty range) and the quantification of the probability of the RF falling outside a certain range, for example, the probability for a negative RF given the time period for the selection of climate drivers.

Figure 14-29 shows PDFs for global-mean forcing, relative to 1750, due to total aerosols, total WMGHGs, and the net of all climate drivers for IPCC AR4 and AR5 as well as two scenarios for 2030 (RCP2.6 and RCP8.5). RFs are presented for AR4 PDFs. All the other results use ERF, but the AR5 WMGHG RF is also shown to illustrate the larger uncertainty in ERF (20%) relative to RF (10%). The mean estimates of ERF and RF for WMGHGs in AR5 are the same, but the much wider PDF for ERF relative to RF is evident.

The change in net forcing between AR4 (1750–2005) and AR5 (1750–2011) results partly from the introduction of ERF, partly from the increase in the WMGHG forcing of 8% due to changes in concentrations between 2005 and 2011, and partly from a wide range of other updates based on a better understanding of processes important for forcing.

The change in the forcing for the two RCPs compared to AR5 (1750–2011) is solely due to the trends in atmospheric composition. The main change in the scenarios are due to aerosols and WMGHGs; the changes to other climate drivers are small (less than 0.1 W m$$^{-2}$$) (Myhre et al. 2015). The weaker aerosol forcing and increased WMGHG forcing enhance the net forcing quite substantially in 2030 (for both RCPs) relative to AR5 (1750–2011) forcing. Furthermore, the weaker contribution from aerosols (with its high uncertainty) and a stronger dominance of WMGHGs (with its relatively low uncertainty) contributes to a smaller uncertainty in the future forcing in both absolute and relative terms. The increase in CO$_2$ in the two scenarios is responsible for between 80% and 100% of the increase in WMGHG forcing and is a strong contributor to the lower uncertainty in the net forcing. The reduction in the magnitude of the aerosol forcing is consistent with recent developments in trends of aerosol abundance (e.g., Paulot et al. 2018).

Since AR5, the WMGHG concentration has increased and there have been updates to RF and the quantification of rapid adjustments (see sections 2 and 3). The upper bar in Fig. 14-30 shows the net anthropogenic ERF from AR5 (which was for the period 1750–2011) with an absolute (5%–95%) uncertainty range of 2.2 W m$^{-2}$ (1.1 to 3.3 W m$^{-2}$). Keeping everything the same as in the upper bar, except updating WMGHG concentrations (to 2018) using the growth rates from NOAA (https://www.esrl.noaa.gov/gmd/aggi/aggi.html) and the methane forcing expression [to include the solar absorption component from Etminan et al. (2016)] gives
the middle bar of Fig. 14-30. The best estimate increases from 2.3 W m$^{-2}$ in AR5 to 2.7 W m$^{-2}$. With a better quantification of rapid adjustments [see Smith et al. (2018) and section 2], the uncertainties in tropospheric rapid adjustment to ERF of CO$_2$ is about 10%, leading to a total ERF uncertainty of 14% compared to the 20% uncertainty assumed in AR5. A large part of the diversity in the ERF of CO$_2$ is likely to be from the instantaneous RF (Soden et al. 2018). The lower bar in Fig. 14-30 combines the contribution of uncertainties in detailed offline calculations (10%) with the 10% uncertainty from climate model–simulated rapid adjustment.

c. Time evolution of forcing

The natural climate drivers from volcanic eruptions and solar irradiance changes have large interannual variations (sections 8 and 9) and because of that, it is difficult to include these in the PDF of climate drivers
for a time period. The time evolution of the natural climate drivers represents their relation to anthropogenic drivers much better than providing forcing over a fixed time period. Hansen et al. (1993a) provided, for the first time, the evolution of various forcing agents and the net RF (Fig. 14-31). Figure 14-31 shows that there is a remarkable similarity between the evolution of net forcing until 2000 between Hansen et al. (1993a) and the estimate in IPCC AR5, given all the new insights since early 1990s. A strengthening in the aerosol forcing especially in the period 1950 to 1980 is further illustrated in Fig. 14-31 as is the strengthening of WMGHG forcing since 1960.

**d. Summary and challenges**

The AR5 result shown in Fig. 14-29 led Myhre et al. (2013) to conclude that “it is certain that the total anthropogenic forcing is positive” strengthening the “extremely likely” wording used in AR4 (Forster et al. 2007). Nevertheless, despite this strengthened language, the AR5 1750–2011 net anthropogenic forcing [2.3 (1.1 to 3.3) W m\(^{-2}\)] indicates that uncertainties remain very large compared to the best-estimate net forcing; this significantly hinders efforts to derive, for example, climate sensitivity given observed temperature changes and inhibits understanding of the effectiveness of proposed mitigation pathways, with consequent impacts on the confidence in the advice given to policymakers.

Challenges to decreasing the spread in the estimated net forcing include (i) ensuring adequate monitoring of changes in concentrations of drivers of forcing and improved understanding of the preindustrial background values, especially for aerosols and ozone, and (ii) improving methods of calculating the forcing given these constituent changes. It is particularly notable that the introduction of ERF in AR5 led to an increase in the uncertainty of the WMGHG forcing. While the relative uncertainty in WMGHG is lower than other components, Myhre et al. (2013) give the absolute 5%–95% uncertainty for the 1750–2011 ERF as 1.14 W m\(^{-2}\); this is only slightly less than the corresponding values for the aerosol–cloud interaction (1.2 W m\(^{-2}\)). Recent results indicate progress in understanding rapid adjustment. Given the expected increasing dominance of WMGHG forcing in coming decades (Fig. 14-29), this indicates the importance of improved estimates of the rapid adjustments in order to reduce the WMGHG ERF uncertainty, as well as efforts to better characterize the other remaining uncertainties in WMGHG forcing. It is also clear that, to date, most efforts on estimating the net forcing have focused on the global mean; further attempts to provide geographical distributions of the net forcing, and the associated confidence levels, would allow additional insights into the drivers of climate change.
11. Emission metrics and radiative efficiencies

As described in previous sections, RF has great utility in quantifying the radiative impact of changes in concentrations of different atmospheric constituents and in assessing their relative importance. RF has also found an important usage in serving climate policy, by enabling methodologies, which are simple to apply in a policy context, to compare the climate impact of emissions of different species; this role is discussed in this section.

a. Background

RF at a given time (e.g., present day), relative to some past time (e.g., preindustrial), is an important indicator of the absolute and relative importance of different drivers of climate change. That forcing depends on the past history of emissions. In the case of CO₂, because of the long persistence times of atmospheric perturbations, emissions over a century ago are still impacting present-day RF; at the other extreme, the present-day RF due (directly) to aviation contrails is mostly the result of contrails formed in the preceding few hours. The influence of the lifetime of perturbations is implicit in, for example, the standard IPCC forcing bars (Fig. 14-2), but this provides little guide to future influence of present-day forcing agents. Figure 14-32 (from Fuglestvedt et al. 2010) illustrates this point for an extreme scenario when a selection of emissions from the transport sector are instantaneously reduced to zero. The forcing due to CO₂ emissions persists for centuries, while most of the forcing due to short-lived species reduces to zero within a few weeks.

From a policy perspective, the explicit consideration of these time scales is important, for example, in assessing the future impact of present-day emissions. In addition, in the context of climate agreements, which cover emissions of a range of different gases (sometimes called multigas agreements), it is necessary to place the climate impact of different emissions on a common scale. Via the application of a climate emission metric (henceforth “metric”), it is (at least in principle) possible to aggregate all emissions into a single “CO₂ equivalent” value.

There are many aspects and choices, some contentious, to consider in the design and application of such metrics, and there is no agreed metric that is suitable for all purposes. The issues have been extensively discussed in various reviews and assessments and has continued through all IPCC ARs (e.g., Fuglestvedt et al. 2010; Myhre et al. 2013). Some of the issues extend beyond physical science and reflect policy choices on, for example, the appropriate time scales and the extent to which the chosen metric serves the aims of a particular policy. Here, the focus will be on the role of RF in computing these metrics.

b. The global warming potential

By way of illustration, one metric, the GWP, is considered briefly here; it is the most widely used metric in international policymaking. The GWP was presented in FAR, based on rather few precursor studies (Derwent 1990; Fisher et al. 1990; Lashof and Ahuja 1990; Rodhe 1990) and has been assessed in all IPCC assessments since then. The GWP is, for example, used by parties to the 1997 Kyoto Protocol to the United Nations Framework Convention on Climate Change (UNFCCC) to place emissions of more than 20 gases on a common CO₂-equivalent scale and in the UNFCCC’s 2015 Paris Agreement. It is also used within the 2016 Kigali Amendment to the UN’s Montreal Protocol on Substances that Deplete the Ozone Layer to place targets on emissions of many hydrofluorocarbons.

The GWP measures the time-integrated RF of a pulse emission of a unit mass of a climate forcing agent (or its precursor) relative to the time-integrated RF of a pulse emission of a unit mass of CO₂. For each agent, it is necessary to know the radiative efficiency (i.e., the RF per molecule or per kilogram) and its lifetime, which determines the decay of the pulse after emission. In addition, indirect forcings resulting from that emission should be incorporated; one example is the impact of methane emissions on ozone, stratospheric water vapor, and CO₂. In almost all policy applications of the GWP, the integration is performed over 100 years (the “time horizon”) and denoted GWP(100). However, there is no compelling scientific reason for that choice, and the perceived importance of emissions of a gas (i.e., their contribution to CO₂-equivalent emissions) can depend markedly on that choice, especially for short-lived species. For example, for methane, AR5 (Myhre et al. 2013) reports values of GWP(20) and GWP(100) of 84 and 28, respectively.

All IPCC assessments have presented values of GWPs for a range of gases [more than 200 are included in Myhre et al. (2013)] and have discussed scientific issues in determining the input parameters. The reported values of GWP(100) for some species have varied quite strongly over time. For example, the GWP(100) for methane has increased from 21 to 28 between FAR (Shine et al. 1990) to AR5 (Myhre et al. 2013) as a result of changes in the recommended values of methane’s radiative efficiency, lifetime, and indirect effects, as well as changes in the radiative efficiency and lifetime of CO₂. Advances in understanding since AR5 [e.g., the effect of methane’s near-infrared absorption bands (section 3)], and the incorporation of the influence of carbon–climate feedbacks (e.g., Gasser et al. 2017; Sterner and Johansson...
could see the recommended GWP(100) value for methane change significantly in the future.

c. Radiative efficiency

Radiative efficiency (RE) is a key input to all the main emission metrics that have been proposed to support multigas agreements. By convention, IPCC assessments have computed the RE for a small perturbation to present-day concentrations for the more major greenhouse gases (CO₂, CH₄, N₂O) for which forcing does not increase linearly with concentration; for more minor species, present in sub-ppbv concentrations, the forcing is assumed to be linear in concentration. RE calculations need to be representative of global-average conditions including the effects of clouds (e.g., Myhre and Stordal 1997). Irrespective of its use in metrics, the RE gives insights into the role of different gases. For CO₂, Myhre et al. (2013) report a value of $1.37 \times 10^{-5}$ W m⁻² ppb⁻¹ (Table 14-2). The RE of methane is 26 times higher, and nitrous oxide’s is 220 times higher. Halocarbons are often greater than 10000 times more effective, per molecule. There are multiple reasons for these differences in RE (e.g., Shine 1991). These include the fundamental spectroscopic intensity of each molecule, which is determined by the probabilities of vibration-rotation transitions, the wavelengths of absorption features relative to the Planck function at typical atmospheric temperatures, the preexisting atmospheric

![Figure 14.32: Consequences for RF of different agents and associated temperature response over time from an assumed scenario when a selection of emissions from the transport sector is instantaneously reduced to zero in 2000. From Fuglestvedt et al. (2010); reprinted with permission from Elsevier.](image-url)
concentrations of the molecule (as RE decreases with concentration), and overlap with absorption features of other atmospheric gases (notably water vapor and CO₂). Hodnebrog et al. (2013) attempted to characterize the sources of uncertainty in calculating REs, focusing particularly on the halocarbons. The overall conclusion was that REs were accurate to within 15% for longer-lived gases and to within about 25% for the shorter-lived gases. Compilations of halocarbon REs and the associated emission metrics in many earlier IPCC assessments drew values from different sources that used various techniques to compute the forcing. This inhibited a reliable comparison of the RE of different gases.

AR5 (Myhre et al. 2013), using the Hodnebrog et al. (2013) calculations, tried to enhance the consistency between gases by adopting a single method for calculating the RE.

To date, REs have mostly been computed using RF (i.e., accounting for stratospheric temperature adjustment in some way), rather than using ERFs for several reasons. First, computing ERFs for such large numbers of gases using GCMs would be a formidable task; in the future, a simpler generic framework for estimating rapid adjustments could be developed if it was shown to be applicable to a wide range of gases. Second, GCM radiation codes do not have the spectral resolution that is necessary for reliable RE calculations for gases with generally quite narrow spectral features. Finally, because of the noise inherent in GCM calculations of ERF, estimation of REs for sub-ppbv concentrations (and hence radiative forcings below about 0.2 W m⁻²) (Forster et al. 2016) would be difficult; while artificially high perturbations could be imposed in GCMs, this would raise questions about the applicability of the results to more realistic concentrations found, or likely to be found, in the atmosphere.

12. Climate response to radiative forcings

Radiative forcing is an essential component in understanding historical changes in climate and their attribution to both natural and anthropogenic causes. Attribution of anthropogenic climate change requires both models and observations and consists of three steps: (i) detecting a change in climate; (ii) establishing that this change is consistent with the expected response to the estimated anthropogenic net forcing; and (iii) establishing that this change cannot be explained by other mechanisms, such as internal variability or natural forcings. Without quantitative information on the magnitude, spatial structure, and temporal evolution of both natural and anthropogenic radiative forcings, understanding the causes of historical changes in climate, from glacial–interglacial periods to the change in climate over the past century, would not be possible.

As an illustration of this process, Fig. 14-33 (left column) compares the observed and model-simulated global-mean temperature anomalies from 1860 to 2012 with that predicted from a multimodel ensemble of coupled ocean–atmosphere models from CMIP3 (gray lines) and CMIP5 (yellow lines) that are integrated under three different forcing scenarios: historical natural and anthropogenic forcings (top), natural forcings only (middle), and anthropogenic greenhouse gas forcings only (bottom). The right columns show the corresponding ERF for each of these forcing scenarios derived from the CMIP5 simulations following Forster et al. (2013).

This set of simulations highlight the basis for the attribution of recent warming to human-influenced activities. The simulations with natural forcing alone (center-left panel) are unable to explain the rise in global-mean temperature from the mid-twentieth century to present. However, when natural and anthropogenic forcings are combined (top-left panel), the model simulations are in excellent agreement with the observed temperature anomalies. Contrasting these simulations with the GHG-only simulations (bottom panels) serves to illustrate how aerosols have partially offset a significant fraction of the GHG warming; without aerosols forcing, the present-day global-mean temperatures would have increased another ~0.5 K. Since anthropogenic aerosols occur primarily in the Northern Hemisphere, these differences also have a distinct signature on the hemispheric contrast in temperature and precipitation as
discussed below. Thus, accurate knowledge of both GHG and aerosol forcings is critical to understanding past and projecting future changes in climate, both globally and regionally.

Early attribution studies focused primarily on the changes in global-mean surface temperature (e.g., Wigley and Raper 1990; Stouffer et al. 1994). These evolved into more complex studies that considered the
spatial pattern or “fingerprints” of climate change that were predicted by climate models and used more elaborate statistical methods to search for their presence in the observed record (Santer et al. 1996b). The fingerprints of anthropogenic climate change were expanded beyond simply the regional pattern of surface temperature to consider the vertical structure of temperature change in the atmosphere and oceans, temperature extremes, and the diurnal and seasonal cycles of temperature, as well as other variables such as precipitation and sea level pressure (Santer et al. 1995; Hegerl et al. 1996, 1997, 2006; Stott et al. 2000, 2004; Vecchi et al. 2006; Meehl et al. 2009a,b; Min et al. 2011; Zhang et al. 2007). While each of these quantities have a defined response to anthropogenic forcing, the use of multivariate changes enhances the confidence in the attribution process. For example, both solar and GHG forcings can warm the surface; however, they have very different impacts on the diurnal temperature cycle, precipitation change, and the temperature response in the stratosphere. Across all of these quantities, the observed changes are only explained through the inclusion of anthropogenic forcing. Thus, the concept of radiative forcing is fundamental to methods used to identify a human influence on climate.

The spatial structure of forcing and response

As noted in sections 2 and 5, anthropogenic forcing from well-mixed GHGs and aerosols over the historical period have distinctly different spatial patterns—the former being positive and more spatially uniform, whereas the latter is a net negative and largely confined to the Northern Hemisphere. As noted above, this hemispheric asymmetry in the pattern of forcing has significant impacts on both the detection and attribution of anthropogenic climate change (Bindoff et al. 2013) as well as the transient climate response to forcing in both historical simulations and future projections (Shindell 2014).

For example, rainfall observations reveal a coherent southward shift of the tropical rain belt over the latter half of the twentieth century, which has been associated with severe droughts over the Sahel and portions of South America (Folland et al. 1986; Allen et al. 2015). Observations in the last few decades of the twentieth century have revealed a weakening of the monsoon (Ramanathan et al. 2005; Chung and Ramanathan 2006; Lau et al. 2006). This tropical precipitation shift has been linked to changes in the cross-equatorial energy transport driven by the interhemispheric contrast in the warming of the surface temperatures that arise from the combination of a spatially inhomogeneous shortwave forcing by aerosols and a relatively more homogeneous longwave forcing by WMGHGs (Kang et al. 2008). This type of a spatial structure in the radiative forcing leads to a decrease of the shortwave flux, primarily felt at the surface in the Northern Hemisphere (e.g., Chen and Ramaswamy 1996; IPCC 2007), and an increase in the heating of the atmosphere and surface globally due to the longwave effects of the WMGHGs (e.g., Ramanathan 1981). The model-simulated response to this pattern of hemispheric asymmetry in the forcing enhances the temperature gradient, induces a change in the meridional circulation, and intensifies the heat and moisture exchange between the two hemispheres across the equator, effectively moving the tropical precipitation belt toward the warmer hemisphere (Ramaswamy and Chen 1997).

Subsequent studies with improved models have affirmed this characteristic for realistic aerosol and WMGH forcing (Rotstayn and Lohmann 2002; Ming and Ramaswamy 2009; Hwang et al. 2013; Shindell et al. 2015; Wang 2015).

Both aerosols and WMGHG influence the spatial pattern of warming. Anthropogenic aerosols introduce hemispheric asymmetry by preferentially cooling the Northern Hemisphere, while WMGHG warms the Northern Hemisphere more than the Southern Hemisphere due to the differing thermal inertia between the two hemispheres (Friedman et al. 2013). When forced by historical changes in both anthropogenic aerosols and greenhouse gases, the cooling effect of the aerosols dominates and climate models simulate less warming in the Northern Hemisphere compared to the Southern Hemisphere.

Aerosols induce an anomalous meridional circulation, which reduces the ascent in the northern tropics and opposes the local Hadley circulation (Ming and Ramaswamy 2011). The hemispheric temperature gradient initiated reduces the southward cross-equatorial energy transport accompanying the weakening of the Hadley circulation (Bollasina et al. 2011; Hwang et al. 2013; Soden and Chung 2017). This change in the large-scale atmospheric circulation, in turn, drives meridional shifts in the tropical precipitation bands (Allen et al. 2015), most notably a southward shift in tropical precipitation compared to simulations in which only GHG or natural forcings are included. The aerosol-induced effect on the cross-equatorial circulation and precipitation changes emerges as a strong spatial feature vis-à-vis WMGHG effects. In addition, the poleward transport of heat in both atmosphere and ocean are affected by aerosols differently than in the case of the WMGHGs (Ocko et al. 2014). Polson et al. (2014) note a correlation between the trend of precipitation in the last few decades of the twentieth century and sulfur emissions. Recent work has also highlighted the role of aerosol–cloud interactions that induce adjustments in cloud interactions in amplifying this response. Aerosol forcing is found to
induce secondary changes in the model-simulated cloud radiative properties through both microphysical (Chung and Soden 2017) and large-scale dynamical changes. These changes act to further increase the hemispheric contrast in forcing, thereby amplifying both the circulation and precipitation changes. However, the broad spatial pattern of the climate response to aerosols and well-mixed GHG forcings is more similar than would be expected given the differing geographic distributions of their emissions and forcings (e.g., Boer and Yu 2003; Levy et al. 2013; Xie et al. 2013).

Boer and Yu (2003) showed that the CO$_2$ forcing spatial pattern (with maxima in the subtropics and a minima at high latitudes) explains very little of the surface temperature response pattern (generally increasing with latitude). While the aerosol forcing pattern explained more of the aerosol response pattern, the correlation remained small; indeed, the CO$_2$ and aerosol surface temperature responses were better correlated with each other than with their “parent” forcing pattern. Recent studies suggest that atmospheric feedbacks to the different patterns of forcing serve to homogenize the radiative perturbations, resulting in a more spatially similar pattern of response for both temperature and precipitation (Ganguly et al. 2012; Xie et al. 2013; Hill et al. 2015; Tian et al. 2017; Persad et al. 2017). The rapid atmosphere-only adjustments to the forcing have been found to be particularly effective at homogenizing the response to aerosols and GHG forcings.

Another spatial dimension to be considered is the vertical. An important parameter to consider in the presence of WMGHG, tropospheric aerosol, and stratospheric ozone forcings is the manner of changes in the vertical structure of the thermal profile and comparison of model simulations with observations (Santer et al. 1996a). In the troposphere, there is a warming of the surface and troposphere primarily due to the combined effect of the positive forcings primarily by WMGHGs (with smaller contributions from short-lived gases and tropospheric ozone) and negative forcing by tropospheric aerosols (Mitchell et al. 1990). In the stratosphere, there is a reduction of temperature due to loss of stratospheric ozone and increases in CO$_2$, tropospheric ozone, and stratospheric water vapor (Shine 1991; Hansen et al. 1995; Ramaswamy et al. 1996; Forster and Shine 1997). The confirmation of the model simulations by observed temperature trends, from the troposphere to the stratosphere and the temporal evolution of the vertical profile of temperature changes (e.g., Santer et al. 1996a; Schwarzkopf and Ramaswamy 2008), has helped affirm our knowledge of the radiative forcing agents, the radiative perturbations they exert on the atmosphere and surface, and the manner of their influences on the spatial dimensions of the climate system (Hansen et al. 1995; Stott et al. 2004; Santer et al. 2005; Ramaswamy et al. 2006a,b; Schwarzkopf and Ramaswamy 2008; Fu et al. 2011).

13. Solar and terrestrial radiation management

a. Introduction

The growth in the understanding and quantification of RF, its consequences for climate change, and the attribution of observed phenomena to forcings, as described in the prior sections, has initiated conceptual thinking on potential management of the forcing(s) to mitigate climate change. The principal pursuit is on how the long-term WMGHG forcings can be partially offset or mitigated. The idea of solar and terrestrial radiation management (STRM), sometimes also referred to as geoengineering, has over the last decade become one of the research areas in which the concept of radiative forcing has been applied. But the idea is by no means new. As early as the 1960s, Budyko wrote about “the possibility of implementing in future some projects of active influence on the climate” (Budyko 1969). Ironically, Budyko discussed this possibility as a way to prevent a new ice age, while the current discussion is focused chiefly on the possibility of intentionally imposing a negative radiative forcing on the Earth system and thus introducing a cooling tendency. Specifically, current research mainly explores the potential of using a few different STRM strategies to counter some or all of the warming from increasing greenhouse gas concentrations in the atmosphere. While the first research on this controversial topic began as early as the 1980s (Keith and Dowlatabadi 1992), it was not until Crutzen (2006) that extensive research on various types of solar and terrestrial radiation management began. The field represents one of the most recent examples of how the radiative forcing concept can be useful for a range of research topics. Several comprehensive reviews and assessments of this literature have been written (Royal Society 2009; Caldeira et al. 2013; Ocean Studies Board 2015). They all share the conclusion that more research on the risks and benefits of STRM is urgently needed.

To date, research on the topic has been carried out almost exclusively with numerical models, and assessment has been based on the modeling activity organized by the Geoengineering Model Intercomparison Project (GeoMIP) (Kravitz et al. 2013b). In the following, we review the major findings that have emerged from numerical modeling to date, with a focus on the viability and climate response of the three types of STRM that have gained the most attention: stratospheric aerosol injection (SAI), marine sky brightening (MSB), and
cirrus cloud thinning (CCT). While the former two strategies fall in the category of solar radiation management (SRM), the latter belongs to what can be labeled as terrestrial radiation management (TRM). Viability in this context refers to the likelihood that the STRM mechanism in question can produce negative forcings of a magnitude sufficient to counter a considerable proportion of anthropogenic greenhouse gas forcing (currently at approximately 3 \( \text{W m}^{-2} \); see section 3).

Here, we focus exclusively on the physical science related to STRM and thus review the literature to date and identify important knowledge gaps related to that aspect only. We do not discuss here any aspect of potential implementation concepts nor factors underlying them.

b. SRM

In the following, studies on the two types of SRM that have so far received the most attention are discussed [sections 13b(1) and 13b(2)]. SRM strategies that have been proposed but received limited attention thus far, including surface albedo enhancements and the introduction of mirrors in space, will not be reviewed here. A review of literature on the climate response to SRM, with a focus on changes to the hydrological cycle, is presented in section 13c.

1) SAI

Extensive research focused on the major volcanic eruptions of the last century have clearly demonstrated the conceptual viability of stratospheric aerosol injection as an SRM strategy (see section 9). However, reported forcings from GCM studies of SAI over the last decade span a wide range, with differences that can mainly be attributed to differing assumptions regarding injection location and height, injection rates, and aerosol sizes; estimates of RF range from \(-8.5 \) to \(-1.2 \text{ Wm}^{-2} \); and the associated global-mean cooling ranges from 0.5 to 3.2 \( \text{K} \) (e.g., Rasch et al. 2008; Robock et al. 2008a; Jones et al. 2010; Berdahl et al. 2014; Crook et al. 2015; Niemeier and Timmreck 2015). Thus, even the most conservative estimates suggest that SAI likely has the potential to offset an appreciable proportion of anthropogenic GHG forcing to date.

Most studies to date have considered stratospheric injection of \( \text{SO}_2 \), which in many respects mimic the climate impact of major volcanic eruptions. The advantage of this is that the global-mean response of the climate system to large injections of sulfur into the stratosphere is well understood, largely due to studies of the 1991 Mount Pinatubo eruption and other well-observed volcanic eruptions (see section 9). However, the spatial responses due to the volcanic aerosol perturbations are not well understood, and there are knowledge gaps with respect to the climate feedback processes operating on the time scale of duration of the stratospheric particles (e.g., cloud interactions, air–sea interactions). Further, the relationship between the sulfur injection rates and the resulting forcing remains poorly understood and is expected to be nonlinear. The nonlinearity with increasing emission rates arises in part because of the corresponding increase in aerosol size, which decreases aerosol lifetime and scattering efficiency. As a result, studies with prescribed aerosol sizes, regardless of emission rate, may overestimate forcing, especially for high emission rates. A major uncertainty in this respect is the aerosol coagulation rate in a freshly injected plume under stratospheric conditions. Solid materials have also been considered for the purpose of SAI (Weisenstein et al. 2015; Jones et al. 2016b) and could considerably reduce the stratospheric ozone destruction that follows from injection of sulfur in the stratosphere. But apart from a few isolated studies, the associated forcing is largely unexplored.Irrespective of the injected material, there is broad agreement that large and negative forcings are achievable through SAI but considerably less agreement when it comes to the climate response beyond the intended global-mean cooling, which will be discussed further in section 13c.

2) MSB

While the very first STRM studies focused on SAI, research to explore the alternative strategy of marine cloud brightening (MCB) also emerged almost three decades ago. Notably, Slingo (1990) demonstrated how even modest perturbations of cloud properties could induce strong radiative forcings. This understanding triggered the idea of intentionally perturbing marine stratocumulus clouds for the purpose of cooling the climate (Latham 1990). Follow-up research based on satellite observations has since identified the most susceptible regions for the purpose of MCB as the sub-tropical regions off the west coasts of the major continents (Alterskjær et al. 2012). By injecting cloud condensation nuclei into the marine boundary layer of these regions, cloud albedo could be increased through the creation of artificial ship tracks, through the mechanisms discussed in section 5. Hill and Ming (2012) employed a coupled-mixed-layer ocean–atmosphere GCM to conduct marine stratocumulus brightening experiments and found that over half of the radiative cooling is due to scattering of solar radiation by the added sea salt aerosols, while the rest arose from enhancement of the local cloud albedo. This finding was recently supported by Ahlm et al. (2017) and suggests that the term marine sky brightening is more appropriate
for this STRM strategy, as the forcing would manifest both under clear and cloudy skies.

While no studies of SAI to date have produced positive forcings, Alterskjær and Kristjánsson (2013) reported from a GCM study that, depending on the size of the injected particles (usually assumed to be sea salt), the competition for the available water vapor could actually lead to a reduction in cloud droplet number concentrations, leading to a positive forcing. The opposite of the desired brightening effect has also been reported from studies using large-eddy simulations, depending on the background aerosol concentration and humidity (Wang et al. 2011). Therefore, despite GCM studies supporting the conceptual viability of MCB as an STRM strategy, important knowledge gaps remain. Notably, many of the early GCM studies on MCB, which produced very large negative radiative forcings, simply prescribed an increase in cloud droplet number in marine clouds and thus could not capture buffering effects like the one described above.

c. Climate response to SRM

Beyond the global-mean cooling that SRM strategies are designed to produce, the following common features of the simulated climate response to SRM have been identified:

(i) SRM naturally yields negative forcings that maximize in the tropics, where insolation is strong. Consequently, SRM tends to cool the low latitudes more than it cools the polar regions, and for a global-mean cancellation of a given GHG forcing, it will generally produce overcooling in the tropics while only partly cancelling the high-latitude GHG warming. This would in turn reduce the equator-to-pole temperature gradient and generate atmospheric circulation changes.

(ii) Along with the cooling comes a reduction in global-mean precipitation, as expected. There is consensus among GCMs that a complete cancellation of GHG warming by SRM in the global mean will lead to an overcompensation in global-mean precipitation. This net precipitation reduction is a consequence of the relative changes to the surface and atmospheric radiation budgets in a climate with both increased GHG forcing and SRM (Bala et al. 2008). It may therefore prove more beneficial to only partly compensate for the increase in GHG forcing (MacMartin et al. 2013).

(iii) A novel finding is that the reduction of shortwave radiation to the surface that would result from SRM may have detrimental effects on agriculture, in contrast to what has been reported in previous studies (Proctor et al. 2018). This finding warrants further investigation.

d. TRM

Only one TRM method has so far received sufficient research attention to merit review here, namely, cirrus cloud thinning. In the following, the research on CCT and the associated climate response are reviewed.

1) CCT

Taking advantage of the fact that most cirrus clouds exert a net positive cloud radiative effect, this TRM strategy proposes to reduce cirrus cloud coverage and lifetime and thus generate a negative forcing. The mechanism by which this could be achieved is through seeding of the cold (<40°C) upper troposphere with ice nucleating particles (INPs), which would allow ice crystals to nucleate and grow even for very slight supersaturations and thus prevent the abundance of ice crystals that results if saturation ratios becomes high enough for ubiquitous tiny solution droplets to freeze spontaneously (~150%). The resulting large ice crystals would sediment out and reduce cirrus cloud cover and upper-tropospheric water vapor, both producing a negative ERF (Mitchell and Finnegan 2009). In theory, this should produce the desired cooling effect, but GCM studies to date produce conflicting results on the matter. The mechanism relies on assumptions about the balance between the dominant ice nucleating mechanisms in the upper troposphere, which is poorly understood in the present. The fact that different studies yield different ERFs should therefore not be surprising.

An additional concern with CCT is the risk of “over-seeding,” that is, injecting too many artificial INPs that ultimately yield higher cirrus ice crystal concentrations than would have occurred in an unperturbed case, and thus a positive ERF (Storelvmo et al. 2013). For this TRM mechanism, the viability is thus still a topic of ongoing research, but recent cloud-resolving modeling results appear to support viability (Gruber et al. 2019) and reinforce the idea that CCT appears to be most promising in the instance of wintertime cirrus clouds at high latitudes.

2) Climate response to TRM

CCT does not reduce global-mean precipitation to the same extent as SRM. The few GCM studies on the climate response to CCT that have been carried out to date suggest that the global-mean hydrological sensitivity (i.e., the precipitation change for a unit temperature change) is similar for CCT and GHG forcing and that simultaneous compensation of temperature and precipitation changes could thus be possible (Kristjánsson
et al. 2015). This appears to hold primarily for mid- and high latitudes (Kristjánsson et al. 2015). The geographical distribution of cooling also more closely mirrors that of GHG warming. With the caveat that the literature on this topic is still limited, it therefore appears that the climate response to TRM is better suited to compensate for GHG warming than SRM. However, as noted above, whether CCT is in fact a viable STRM strategy remains unclear.

3) CONCLUSIONS, KEY UNKNOWNS, OUTLOOK

STRM remains a scientifically pursued but controversial topic, and arguments have been presented in the literature for why even theoretical and modeling studies on the topic should be conducted with caution. The current understanding of the forcings associated with STRM strategies can be summarized as follows:

1) SAI can produce strong negative forcings, but the nonlinear relationship between the injected mass and forcing is poorly understood.
2) MSB through, for example, sea salt injection will very likely achieve negative forcings of the desired magnitude, but there is a nonnegligible chance that forcing of the opposite sign could result under some conditions. Forcing per mass of injected sea salt aerosol is poorly understood and is highly dependent on the injected aerosol size.
3) Ocean albedo modification could generate strong negative forcings but does not seem viable because of the likely interference with ocean ecosystems. A multidisciplinary team of experts is needed to fully address its viability.
4) Proposed land surface albedo modifications are deemed ineffective for the purpose of STRM.
5) Mirrors in space can produce negative forcings of the desired magnitude, but viability depends on the engineering question of whether it is feasible to arrange for sufficient reflective material at the location in space where this would be optimal.
6) In part due to conflicting modeling results, it is still unclear whether CCT can produce sufficiently strong negative forcings to be a viable TRM strategy.
7) Using SAI and MCB to completely cancel global-mean GHG warming would likely cause a net reduction of precipitation relative to preindustrial conditions. A partial compensation of global-mean GHG warming could lead to a lesser effect on precipitation.
8) Should it prove viable as an STRM strategy, CCT appears more suited for an offset of the climate response to increased GHG concentrations.

Common for many of the unknowns related to STRM strategies is that they are related to uncertainties about the unperturbed atmosphere and to processes that are highly relevant for our understanding of the present atmosphere and anthropogenic perturbations to it. STRM research efforts are best directed toward activities that have the dual benefit of increasing understanding both of how past inadvertent and potential future inadvertent forcings affect Earth’s global climate system.

14. Overall summary and challenges

a. Summary

This paper has traced the evolution of the concept of radiative forcing over the past century. It has also described the historical milestones in the scientific community’s understanding of the RF agents, their quantification including the total or net RF, and some important applications stemming from the concept. Beginning with the fundamentals of radiation physics, principally with the developments in the late eighteenth and through the nineteenth and twentieth centuries to the present, that growth has established a powerful framework to quantify the factors that force the climate system by perturbing the shortwave and longwave energy disposition in the Earth system. We have focused on the forcing as defined from preindustrial time (~1750) to present, approximately the mid-2010s. [We note that IPCC AR6 will be completing a major assessment in 2021, which will represent a major full update since the IPCC (2013) assessment.] We have treated the fundamental developments by considering the major inflection points of scientific advances, more particularly as they relate to the quantification of the estimates together with the uncertainties. For some agents, such periods are well marked by times prior to approximately the 1950s (say, before the International Geophysical Year, 1957), between the 1950s and the advent of the satellite era of global measurements (1979), and the onset of the major international assessments since the mid-1980s. We have highlighted especially those points in time when the literature saw more robustness added to the knowledge especially with regards to quantification. The stages in growth evolved differently for the different forcing agents, with the complexity of some still hindering a rapidity of growth in confidence in the quantitative estimates (e.g., aerosol indirect forcing).

The forcing concept has been conceptualized and applied with the intent to define a metric that would be helpful in providing a first-order estimate of the global-mean climate impact and most specifically the effect on surface temperature. Perhaps the most important application has been the use of the RF estimates to comparatively estimate the climate responses due to different agents, both for policy decisions concerning mitigation and adaptation and for the scientific understanding of the relative importance of the
Forcing agents. In recent years, the forcing concept has been extended to investigate the comparative impact of forcings on changes in circulation patterns, including consideration of both anthropogenic and natural forcings, and taking into account the internal variability of the climate system. The RF concept has also been used to formulate simple metrics for the warming potential of various agents and for ideas in radiation management.

There are several limitations encountered with both the RF and the more recent ERF concepts. These include accounting for the growing recognition of the complexity of some of the forcings, inadequacies in characterizing and narrowing uncertainties in the determination of the forcings, uncertainties in precursor parameters to determine the forcing (e.g., preindustrial emissions), reliance on numerical models for estimates of forcing, and difficulties in achieving consistency across different numerical models and their estimates. The above issues in turn affect the synthesis of the estimate and uncertainty of the total RF of climate.

RF and ERF remain theoretical concepts well suited to computational estimates and computer modeling of the climate system including the agents that drive climate change. However, there are shortcomings. These include the inability to observe/measure/quantify parameters of relevance in forcing estimates, the inability to monitor on a continuous basis key radiative flux and associated parameters to a high degree of accuracy, difficulties in verifying the theoretically formulated radiative forcing against practically measurable observations since the latter are not rigorously able to measure the changes in the state of the system without feedbacks, and theoretical (and laboratory measurement) gaps inhibiting the knowledge of the processes that lead to the agent’s forcing of climate.

Weaknesses in the application of the concept include propagation of the uncertainties in different factors leading to the forcing estimate. This in turn affects the evaluation of the feedbacks and response due to the forcing. It also impacts linkages to detection and attribution of climate change, ability to link observed phenomena (e.g., extreme events) to forcing of the climate system, and narrowing uncertainties in climate projections.

It is important to note that the RF concept has been more than just an arbitrary or an academic formulation. It has gone beyond just a routine definition to express a metric concerning the climate driving/forcing agent. Despite its limitations, there have been remarkable successes with this metric in understanding the way global climate responds to different forcings. Many findings have been demonstrated to be highly useful: for example, links to emissions/sources/precursors and relative effectiveness of the different forcers, the ability to intercompare models to quantify the changes in climate parameters, and policy applications (e.g., GWPs).

Additional relevance with regards to climate impacts are the following points:

- The RF concept has found usefulness in the policy context as a precursor to global warming potential—paralleling ozone depletion potentials in the ozone loss context. It has become possible to link RFs of forcings to their influences on surface temperature, thus allowing cross comparison of the climate change effectiveness of different forcers. It has also enabled thinking to develop along the lines of mitigation of anthropogenic greenhouse gas effects such as in solar radiation management (Royal Society 2009).
- RF has become intertwined in the linkage from emissions to responses. A degree of robustness has been obtained using 3D global climate models to connect RF and responses in a simple manner. This started with global means but has also included regional-based quantitative estimates in the case of some forcers. The variable in question has been principally surface temperature, but extensions have been made to other climate parameters through physical connections (hydrologic cycle and precipitation, sea level rise).
- Forcing at the TOA/tropopause, that at the surface, and their physical relevance have been identified with changes in parameters describing the physical climate change. While surface flux change does not relate easily to surface temperature change, it can be linked to hydrological cycle and precipitation. (e.g., Asian monsoon and hemispheric/global spin-down of precipitation caused by aerosols).
- Significance of the question: What is Earth’s climate sensitivity to radiative forcings? Estimating climate sensitivity from observed temperature changes depends crucially on our knowledge of RF. If net anthropogenic RF over the past century has been the result of a significant offset of the positive forcings by negative forcings, that would suggest a strong climate sensitivity. But if positive forcing has been the dominant type, then the system is relatively less sensitive. The dependence on an accurate RF to obtain a good estimate of climate sensitivity crucially determines in turn how severe climate impacts are likely to be under the influence of increased GHG emissions in the twenty-first century estimates. Accompanying this task should be the sustained monitoring of changes in aerosol and related properties, for example, aerosol optical depths, vertical profile, clear- and all-sky spectral and total radiative fluxes.
- Characterizing and, to the extent possible, narrowing the uncertainties in the preindustrial state especially
for short-lived climate forcers (e.g., ozone, aerosols). Nonlinearities and interactions between forcings and its invariable relation to feedbacks (e.g., dependencies on cloud, water vapor, surface albedo).

- Continue to explore the potential for direct observation of radiative forcing (e.g., stratospheric aerosols in the aftermath of a volcanic eruption, solar irradiance, spectrally resolved TOA, and surface observations).
- As the ERF concept advances, clarifying and documenting the nuances that differentiate this from RF (e.g., the uncertainties arising due to different treatments of physical processes in different models, efficacy factors).

**b. Important research challenges in the coming decades**

**IMPROVING THE ACCURACY OF THE FORCING ESTIMATE**

The principal challenge in arriving at a definitive estimate of the net forcing since preindustrial times is made starkly evident by the IPCC assessments. The last IPCC assessment report (2013) conclusion that the 1750–2011 net anthropogenic forcing (best estimate of $2.3 \text{ W m}^{-2}$, with a range from 1.1 to $3.3 \text{ W m}^{-2}$) indicates that the remaining uncertainties are very large compared to the best-estimate net forcing; this significantly hinders efforts to derive, for example, climate sensitivity given observed temperature changes and inhibits understanding of the effectiveness of proposed mitigation pathways, with knock-on impacts on the confidence in the advice that can be given to policymakers:

- Ensuring GCM radiation codes faithfully represent radiative processes—with linkages to available observations. One particularly demanding task is to reduce the undesirable spread in the CO$_2$ forcing especially for the purposes of computing accurate climate responses in climate models. The World Climate Research Programme Radiative Forcing Model Intercomparison Projects have enabled the community to calibrate models against robust reference calculations. The latest venture (Pincus et al. 2016) is expected to sustain the momentum and push the frontiers further forward.
- Deployment and utilization of observations from multiplicity of platforms for characterizing fully the four-dimensional distribution of forcing agents and their time evolution, especially for natural forcers such as solar irradiance and volcanic aerosols, and for short-lived climate forcers such as aerosols and related cloud microphysics and ozone.
- A better process understanding of aerosol forcing and in particular aerosol–cloud interactions, which has been a main contributor to uncertainty in anthropogenic forcing.
- Computational and observational determinations that need to be carried out include the following:
  - Taking advantage of the rapid advances in the past few years in computational architectures, algorithmic formalisms, and computing capacities. Recent advances include the facilitation of machine learning and neural networks (e.g., Krasnopolsky et al. 2005), performing line-by-line benchmark computations over the entire global scale of a model (e.g., A. L. Jones et al. 2017).
  - Processes and model parameterizations that enable translating the radiative forcing to climate change, including the physical, chemical, and dynamical responses to the forcing, and the modulation of extant circulation patterns throughout the integrated system.
  - Continuous monitoring of the agents (concentrations, radiative properties) with the highest possible accuracy and repeatability.
  - Resolution of differences among observed (to the degree feasible) and modeled absolute variations and changes in forcings.
  - Improving the ability of climate models to capture the responses to the natural and anthropogenic forcings and evaluate the resulting changes in climate variables.

It is sobering to realize that RF largely evolved as a theoretical concept, a simple metric to compare effects due to different forcers initially formulated for well-mixed greenhouse gases, and shown to be somewhat easily relatable to simulated surface temperature change. In the past, when running climate model calculations was computationally taxing, RF proved to be an invaluable quantifying capability. With increased computational resources now to run climate responses of single forcings or subsets of forcings or all forcings somewhat inexpensively, is RF or ERF a redundant concept? Recognizing the caveats and deficiencies noted above, and with the newer nuances in the forcing definition (e.g., adoption of ERF as the new forcing metric; section 2), a legitimate question arises—What is the future of the radiative forcing concept?

**c. Grand challenge**

The answer to the above question is that investigations into forcing and estimating it may no longer be independent of considerations of the rest of the climate system. For instance, adjustments (feedbacks) in climate parameters may need to be increasingly considered in order to relate to surface temperature change, which
was the initial quest in determining RFs. RF and ERF remain useful physically based constructs that, despite the fact that they cannot be in general observationally verified, still retain a simple link to climate responses, at least in the context of global-mean surface temperature.

Going forward, the grand challenge lies in viewing RF, and forcing in general, not as a separate entity in the trinity of forcing, feedback, and response, but to be cast into a broader perspective. A picture of this might comprise looking at the Earth system in its broader scope than just temperature and only the physical climate system. Instead, in combining forcing, feedback, and response, we are entering the era of an Earth system challenge that needs to account for feedbacks; for example, ERF departs from RF in allowance for fast feedbacks to be factored into the forcing estimate (section 2). This increase in complexity is manifest for the various forcing agents (sections 3–9). These complexities then add to the body of uncertainties in the net forcing and its applications (sections 10–13), with concomitant impacts on societal adaptation and mitigation planning and measures. There are thus significant facets to the grand challenge regarding the application of RF and ERF in the future and the societal utilization of this science for decision-making.

Radiative forcing in the climate science and climate change context is emerging after a century of exploration and investigations with a firm qualitative sense but still with limitations in its quantitative certainty. There is considerable ground still to be covered in order to achieve a comprehensive resolution of the remaining uncertainties while retaining the simplicity of the concept, and meeting the demand for more accurate quantification of both the agentwise and net anthropogenic forcing.

Acknowledgements. We thank the AMS Centennial Monograph committee for the opportunity to describe the history of radiative forcing. We acknowledge the assistance of Ms. Morina Royer in the preparation and organization of the manuscript. We thank the reviewers and editor for their constructive comments and suggestions. JMH was partly funded under the CLARIFY-2017 Natural Environment Research Council funded proposal (NE/L013797/1). GM has received funding from the Norwegian Research Council project KeyClim (295046).

APPENDIX

Acronym List

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCENT</td>
<td>Atmospheric Composition Change: The European Network of Excellence</td>
</tr>
<tr>
<td>ACCMIP</td>
<td>Atmospheric Chemistry and Climate Model Intercomparison Project</td>
</tr>
<tr>
<td>ACE</td>
<td>Aerosol Characterization Experiment</td>
</tr>
<tr>
<td>ACI</td>
<td>Aerosol–cloud interactions</td>
</tr>
<tr>
<td>ACP</td>
<td>Atmospheric Chemistry and Physics</td>
</tr>
<tr>
<td>ACRIM</td>
<td>Active Cavity Radiometer Irradiance Monitor</td>
</tr>
<tr>
<td>ACRIMSat</td>
<td>Active Cavity Radiometer Irradiance Monitor Satellite</td>
</tr>
<tr>
<td>AeroCom</td>
<td>Aerosol Comparisons between Observations and Models</td>
</tr>
<tr>
<td>AERONET</td>
<td>Aerosol Robotic Network</td>
</tr>
<tr>
<td>AFCRL</td>
<td>Air Force Cambridge Research Laboratory</td>
</tr>
<tr>
<td>AGCM</td>
<td>Atmospheric general circulation model</td>
</tr>
<tr>
<td>AIRS</td>
<td>Atmospheric Infrared Sounder</td>
</tr>
<tr>
<td>AMOC</td>
<td>Atlantic meridional overturning circulation</td>
</tr>
<tr>
<td>AMS</td>
<td>Aerosol Mass Spectrometer</td>
</tr>
<tr>
<td>AOD</td>
<td>Aerosol optical depth</td>
</tr>
<tr>
<td>AR</td>
<td>Assessment Report</td>
</tr>
<tr>
<td>AR4</td>
<td>Fourth Assessment Report</td>
</tr>
<tr>
<td>AR5</td>
<td>Fifth Assessment Report</td>
</tr>
<tr>
<td>AR6</td>
<td>Sixth Assessment Report</td>
</tr>
<tr>
<td>ARM</td>
<td>Atmospheric Radiation Measurement Radiometer</td>
</tr>
<tr>
<td>AerChemMIP</td>
<td>Aerosols and Chemistry Model Intercomparison Project</td>
</tr>
<tr>
<td>BC</td>
<td>Black carbon</td>
</tr>
<tr>
<td>BGC</td>
<td>Biogeochemistry</td>
</tr>
<tr>
<td>C4MIP</td>
<td>Coupled Climate–Carbon Cycle Model Intercomparison Project</td>
</tr>
<tr>
<td>CALIOP</td>
<td>Cloud–Aerosol Lidar with Orthogonal Polarization</td>
</tr>
<tr>
<td>CALIPSO</td>
<td>Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations</td>
</tr>
<tr>
<td>CAM</td>
<td>Community Atmosphere Model</td>
</tr>
<tr>
<td>CATS</td>
<td>Cloud–Aerosol Transport System</td>
</tr>
<tr>
<td>CCM</td>
<td>Chemistry–Climate Model</td>
</tr>
<tr>
<td>CCN</td>
<td>Cloud condensation nuclei</td>
</tr>
<tr>
<td>CCSP</td>
<td>Climate Change Science Program</td>
</tr>
<tr>
<td>CCT</td>
<td>Cirrus cloud thinning</td>
</tr>
<tr>
<td>CDNC</td>
<td>Cloud droplet number concentration</td>
</tr>
<tr>
<td>CDR</td>
<td>Climate Data Record</td>
</tr>
<tr>
<td>CERES</td>
<td>Clouds and the Earth’s Radiant Energy System</td>
</tr>
<tr>
<td>CF₄</td>
<td>Carbon tetrafluoride</td>
</tr>
<tr>
<td>CFC</td>
<td>Chlorofluorocarbon</td>
</tr>
<tr>
<td>CH₄</td>
<td>Methane</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>CICERO</td>
<td>Centre for International Climate and Environmental Research</td>
</tr>
<tr>
<td>CLARIFY</td>
<td>Cloud Aerosol Interaction and Forcing</td>
</tr>
<tr>
<td>CLM4</td>
<td>Community Land Model, version 4</td>
</tr>
<tr>
<td>CM2</td>
<td>GFDL's Coupled Model, version 2</td>
</tr>
<tr>
<td>CMIP</td>
<td>Coupled Model Intercomparison Project</td>
</tr>
<tr>
<td>CMIP3</td>
<td>Coupled Model Intercomparison Project, phase 3</td>
</tr>
<tr>
<td>CMIP5</td>
<td>Coupled Model Intercomparison Project, phase 5</td>
</tr>
<tr>
<td>CMIP6</td>
<td>Coupled Model Intercomparison Project, phase 6</td>
</tr>
<tr>
<td>CO</td>
<td>Carbon monoxide</td>
</tr>
<tr>
<td>CO2</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>CTM</td>
<td>Chemistry transport model</td>
</tr>
<tr>
<td>DMS</td>
<td>Dimethyl sulfide</td>
</tr>
<tr>
<td>DU</td>
<td>Dobson unit</td>
</tr>
<tr>
<td>ECMWF</td>
<td>European Centre for Medium-Range Weather Forecasts</td>
</tr>
<tr>
<td>EESC</td>
<td>Effective equivalent stratospheric chlorine</td>
</tr>
<tr>
<td>ENSO</td>
<td>El Niño–Southern Oscillation</td>
</tr>
<tr>
<td>ERF</td>
<td>Effective radiative forcing</td>
</tr>
<tr>
<td>ESM</td>
<td>Earth system model</td>
</tr>
<tr>
<td>FAR</td>
<td>First Assessment Report</td>
</tr>
<tr>
<td>FDH</td>
<td>Fixed dynamical heating</td>
</tr>
<tr>
<td>FFBC</td>
<td>Fossil fuel black carbon</td>
</tr>
<tr>
<td>FFOC</td>
<td>Fossil fuel organic carbon</td>
</tr>
<tr>
<td>GCM</td>
<td>General circulation model</td>
</tr>
<tr>
<td>GCTMs</td>
<td>Global chemical transport models</td>
</tr>
<tr>
<td>GEISA</td>
<td>Gestion et Etude des Informations Spectroscopiques Atmosphériques</td>
</tr>
<tr>
<td>GFDL</td>
<td>Geophysical Fluid Dynamics Laboratory</td>
</tr>
<tr>
<td>GHG</td>
<td>Greenhouse gas</td>
</tr>
<tr>
<td>GISS</td>
<td>Goddard Institute for Space Studies</td>
</tr>
<tr>
<td>GWP</td>
<td>Global warming potential</td>
</tr>
<tr>
<td>GeoMIP</td>
<td>Geoengineering Model Intercomparison Project</td>
</tr>
<tr>
<td>H2S</td>
<td>Hydrogen sulfide</td>
</tr>
<tr>
<td>H2SO4</td>
<td>Sulfuric acid</td>
</tr>
<tr>
<td>HALOE</td>
<td>Halogen Occultation Experiment</td>
</tr>
<tr>
<td>HCFC</td>
<td>Hydrochlorofluorocarbons</td>
</tr>
<tr>
<td>HFCs</td>
<td>Hydrofluorocarbons</td>
</tr>
<tr>
<td>HIRS</td>
<td>High Resolution Infrared Radiation Sounder</td>
</tr>
<tr>
<td>HITRAN</td>
<td>High-resolution transmission molecular absorption</td>
</tr>
<tr>
<td>HNO3</td>
<td>Nitric acid</td>
</tr>
<tr>
<td>HO2</td>
<td>Peroxy hydroxyl</td>
</tr>
<tr>
<td>HadGEM2</td>
<td>Hadley Centre Global Environment Model, version 2</td>
</tr>
</tbody>
</table>


Humphreys, W. J., 1913: Volcanic dust and other factors in the production of climatic changes, and their possible relation to


Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying climate feedbacks using...


