



Uncertainties and confidence in stratospheric aerosol injection modelling: a systematic literature review

Anni Määttänen, Thibaut Lameille, Carola Klöck, Olivier Boucher, François Ravetta

► To cite this version:

Anni Määttänen, Thibaut Lameille, Carola Klöck, Olivier Boucher, François Ravetta. Uncertainties and confidence in stratospheric aerosol injection modelling: a systematic literature review. Oxford Open Climate Change, 2024, 4 (1), pp.kgae007. <10.1093/oxfclm/kgae007>. <insu-04634987v2>

HAL Id: insu-04634987

<https://insu.hal.science/insu-04634987v2>

Submitted on 15 Oct 2024


HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Uncertainties and confidence in stratospheric aerosol injection modelling: a systematic literature review

Anni Määttänen ^{1,*}, Thibaut Lameille¹, Carola Kloeck², Olivier Boucher³, François Ravetta¹

¹LATMOS/IPSL, Sorbonne Université, UVSQ Université Paris-Saclay, CNRS, Paris, France

²CERI Sciences Po, Paris, France

³Institut Pierre-Simon Laplace, Sorbonne Université/CNRS, Paris, France

*Corresponding author. LATMOS/IPSL, Sorbonne université, UVSQ Université Paris-Saclay, CNRS, Paris, France. Tel: +33144274970, fax: +33180285200.

E-mail: anni.maattanen@latmos.ipsl.fr

Abstract

Model projections performed to evaluate the efficacy and impacts of solar geoengineering interventions, such as Stratospheric Aerosol Injection (SAI), include multiple sources of uncertainty, namely scenario, model, and natural variability uncertainty. It is well accepted that a quantitative uncertainty assessment related to SAI modelling is required to provide robust and policy-relevant information on SAI. This study investigates how and to what extent articles using a climate modelling approach on SAI quantify and communicate uncertainty sources. We conducted a systematic literature review of a sample of 60 peer-reviewed articles in order to (i) analyse whether uncertainties were addressed, and if yes, which methods were used to characterize uncertainties, and (ii) study how the articles communicated assumptions and limits that contribute to the estimation of confidence in the used models and the resulting projections. We present statistics on the uncertainty quantification methods used in the articles and we discuss the vocabulary employed for conveying these uncertainties and model confidence. In the studied article sample, the attention paid to uncertainty estimations in the SAI literature increased with time, and overall, uncertainties were treated using a variety of methods. Model confidence was not always explicitly communicated as the models used are already tested in the literature and their strengths and weaknesses are known to the community although this is often implicit. Our results show that it is currently difficult to perform global, quantitative assessments of uncertainty related to SAI research, in line with recent review reports on solar geoengineering.

Lay Summary

We review a sample of 60 articles on literature studying the so-called solar geoengineering method consisting of injecting aerosols into the stratosphere (above 10–15 km of altitude) using climate models. We focus on how these articles address and quantify the uncertainties related to the model results, and we also study the way the articles justify the use of certain models, the confidence attributed to the models, and the vocabulary used to convey all of these aspects. The articles use a variety of methods to address modelling uncertainties, and the attention paid to uncertainty estimation increases with time. Often the used models are implicitly assumed reliable for the purpose they are used for as they are well known in the field. The solar geoengineering community has recently called for a more systematic treatment of uncertainties that will also enable better estimations of the possible risks of solar geoengineering in future assessment reports.

Keywords: Solar Radiation Management, Stratospheric Aerosol Injections, Uncertainty, Modelling

Introduction

There is currently a broad scientific consensus that the Earth's climate has experienced human-induced changes over the last century. Past, current and future greenhouse gas emissions will continue to impact the Earth's radiative balance and the climate [1]. In this context, climate-related risks¹ for natural and human systems are projected to increase as global temperature increases from pre-industrial conditions to +1.5°C and beyond [2, 3].

The magnitude of these projected risks, the emphasis on the inertia of the climate-carbon system as well as pessimistic views on humanity's ability to mitigate greenhouse gas emissions in a timely fashion [4] have led a small research community to

intensify research activities on solar geoengineering methods. The research aims at evaluating the feasibility, efficacy and specific risks of using these techniques as part of the portfolio of policy options to moderate climate hazards [5–10]. Several review reports on solar radiation management (SRM) research have been published [1, 11–13]. They also highlight the high (but not fully quantified) level of uncertainty involved with such techniques.

'Solar geoengineering' refers to technologies that could be used to intentionally modify the Earth's radiative balance. One of the proposed methods, Stratospheric Aerosol Injection (SAI), consists in injecting aerosol precursors in the stratosphere in order to increase the fraction of sunlight reflected back to space and

¹ In this context, 'climate-related risks' can be understood as potential adverse consequences of climate hazards for human and natural systems, given the circumstances that characterize the vulnerability and exposure of these systems to these hazards [2].

thus inducing a net global cooling at the surface [8, 14–17]. This technology has been conceptualized in analogy with volcanic eruptions, such as the Mount Pinatubo eruption in 1991, which had a discernible (about 0.5 K), relatively short-term (approximately 1 year) cooling effect on the climate [18]. However, volcanic eruptions and intentional injections would not have exactly the same impacts as they are not perfect analogues, differing for instance in terms of stratospheric aerosol size and composition or in the spatio-temporal characteristics of the injection [19–21].

The envisioned scale of SAI deployment is planetary and multidecadal, up to over a century [22]. However, possibilities for obtaining observational data from SAI field experiments are limited to small-scale experiments. Such experiments might help in improving small-scale process descriptions in models and reducing uncertainties related to such processes [10, 23], but they cannot inform us on the large-scale climate impacts of SAI nor do they allow for an experimental validation of climate responses to SAI [24, 25]. Furthermore, recent attempts to proceed to outdoor field testing (e.g. tests of the delivery system for the SPICE project or the SCoPEX project [26]) have faced systematic opposition from parts of the society and have been abandoned. Hence, any discussion on using SAI as part of the portfolio of options to reduce potential future consequences of climate change has so far predominantly relied on climate simulations investigating related stratospheric processes as well as potential climate response and impacts, and it is likely that it will remain so in the foreseeable future. These simulations are performed with numerical climate models that can explore how characteristics (e.g. global surface temperatures) of the modelled climate system would evolve in response to specific forcings (e.g. under greenhouse gas and aerosol injection scenarios). These results can also be used as input data to other types of models (e.g. crop models, hydrological models, etc.) that simulate specific impacts for a particular climate scenario.

Due to the unprecedented nature and vast scale of the potential SAI interventions, related decision-making processes will necessitate a high level of confidence in model projections [10], both those addressing climate change *per se* and those on the impact of SAI on the climate, so that all options and the associated uncertainties can be objectively weighed. However, it is not clear how the necessary ('high') level of confidence for decision-making can and should be defined, but it would require a multidisciplinary approach and a constructive dialogue between scientists and decision-makers [27]. Generally, the absence of a measure of reliability as well as the non-characterization of uncertainties make the usability of scientific studies in a decision-making context more challenging [28]. In this context it is very important that all the underlying assumptions made (e.g. methodological, epistemological) in producing these projections are discussed and uncertainties systematically characterized to provide a measure of the reliability of the projections [29]. Also, explicitly communicating underlying methodological assumptions and characterizing sources of uncertainty are very important in contextualizing the results of these projections [30].

While individual scientific articles are mainly written for peers and not for a general audience, these articles might be also used in a policy-making context and in communication to a general public [29], requiring contextualization in the communication of scientific knowledge outside the scientific community. From this viewpoint, assessing uncertainties on a broad topic to be communicated outside the scientific community requires making a synthesis of the aggregated information from a large spectrum of studies, going beyond the treatment of specific

uncertainties relevant to a single study. A recent systematic literature review focused on uncertainty treatment in modelling and decision-making research articles on the impact of climate change on forests [31]. They identified an inconsistent use of the notion of uncertainty and a lack of its systematic characterization in the reviewed literature, and showed that modelling and decision-making studies addressed different types of uncertainties. The authors concluded that this difference in uncertainty treatments could make communication of research to decision-makers more difficult, and suggested that the scientific literature should address and communicate uncertainties in a clearer and more comprehensive manner, so that it can fully play its role on providing evidence for decision-making [31]. Two other reviews have recently focused on the communication steps following the production of scientific knowledge. One of them performed a qualitative systematic review of studies in behavioural and cognitive sciences on the communication of uncertainties in climate-related scientific findings, focusing on the communication of climate-related information to its user (e.g. the general public) in a decision-making context [32]. The authors identified three sources of uncertainty in their literature sample, focusing on uncertainty in observations of climate change (the two others being deep uncertainty including ignorance and uncertainty on how well climate models reproduce observations). One of the findings was that presenting ranges to respondents of the studies instead of a single numerical value for a climate parameter improved the understanding of the uncertainty [32]. The authors also noted that the terms in which the ranges were described could modify the perception of the respondents, underlining the importance of the vocabulary used in communicating scientific uncertainties [32]. The other review focused specifically on communicating model uncertainties to policymakers [27]. The authors concluded that instead of including all possible uncertainties, the clearest communication is acquired with decision-relevant uncertainties that are related to the policymakers' needs and whose relevance is defined through a dialogue [27]. In addition to the above, communication of results to the general public and decision-makers should also take into account inclusivity, cultural differences and particularities of the different regions of the world [30, 33].

The communication of lines of evidence in the Intergovernmental Panel on Climate Change (IPCC) assessment reports provides a good example of the transition from scientific knowledge for specialists to information targeted to a larger audience. Glossaries and guidance that frame the use of key concepts such as uncertainty and risk were conceived to homogenize practices among and between the IPCC Working Groups [2, 34, 35], and climate scientists were asked to evaluate their confidence in conclusions drawn from climate projections, using these guidelines to express the elements of communication, including, when necessary, a varying degree of expert judgement. In the IPCC assessment reports two metrics are produced: (i) the 'relative degree of certainty' defined as a combination of the degree of agreement and the consistency of evidence, and (ii) if and when possible, estimated likelihood, defined as the probabilistic estimate of a specific occurrence or attributed to a range of outcomes [34]. These two metrics are the results of informed but subjective processes of weighting theory (perceived understanding of the functioning of underlying processes), data, model simulations and expert judgment. The latter is 'only made possible by the large amount of model-based information which resulted from a concerted and timely effort of the climate modeling community' [34]. This means that although individual articles studying climate change might not provide thorough uncertainty estimates, the climate research

community as a whole has produced sufficient information to enable a reliable estimation of these metrics used in the IPCC reports.

In this article, we focus on the production of information by the scientific community that is needed prior to informing a larger audience.

The SRM research community has recently highlighted the need for a better estimation of uncertainties, also in support for decision-making, and provided perspectives on how to proceed [10, 36]. But how has published research so far established and communicated confidence and uncertainties related to modelling studies? The goal of our article is to study the treatment of uncertainties and the attribution of confidence in models in the published SAI modelling literature. We conduct a systematic literature review of a sample of 60 peer-reviewed research articles to identify the used uncertainty characterization methods and the communication of assumptions, limits and results in the SAI literature. In the following, we first describe our methodology (Methodology and terminology section), then we present the results (Results of the systematic literature review section) and discuss our findings (Discussion section), and conclude (Conclusion section).

Methodology and terminology

This section discusses the methodology we employed in this study and provides the context for interpreting common practices identified in the SAI literature and discussing their implications on establishing a degree of confidence. We describe the main sources of uncertainty in climate projections and the methods climate modelers use to investigate these uncertainties. We also discuss factors used in assigning confidence in models and projections.

A Systematic literature review

The questions we try to answer with the systematic literature review method address two topics:

- Confidence and uncertainties in SAI modelling studies:
 - were uncertainties addressed or not?
 - which types of uncertainty were addressed, and with which methods?
 - how were model performance and confidence estimated?
- Communication of confidence in SAI modelling studies:
 - how was confidence communicated, and with which terminology?

We conducted a systematic literature review of a sample of 60 peer-reviewed articles that study different SAI strategies through a climate modelling approach. We used the systematic literature review to reveal common practices for communicating uncertainties and underlying assumptions in the SAI literature. The methodology of systematic literature review is described in the literature [37, 38].

We used the Scopus database to search for peer-reviewed research articles published in English with the following search terms: (TAK = Title-Abstract-Key Words) TAK (solar AND geoengineering) OR TAK (strato* AND aerosol AND geoengineering) OR TAK (strato* AND aerosol AND injection) OR TAK (solar AND radiation AND management AND geoengineering). This research string takes into account that different terms are commonly used when referring to these techniques (e.g. solar geoengineering, solar radiation management, stratospheric aerosol injection). The search string we used might exclude some articles not using the term 'geoengineering' at all, but the search substring TITLE-ABS-KEY

(strato* AND aerosol AND injection) ensures that SAI articles are in any case included. We did not limit the temporal range of our search so that we sample the full corpus of SAI publications. The search in the Scopus database was conducted on 15 February 2021. These search terms yielded 656 articles. Based on the abstracts, we excluded off-topic articles (e.g. those only mentioning stratospheric aerosol injections but not studying them, articles on other solar geoengineering methods) and only kept primary research studies using a climate modelling approach, also including the ones mimicking the impact of stratospheric aerosol injections with a reduction of the solar constant. With these inclusion/exclusion criteria, the sample was reduced to 193 articles. From the remaining 193 articles, 60 were randomly selected in order to have a sample size that allowed an in-depth text analysis. Information collected in this analysis was coded in an Excel sheet (provided as a [Supplementary File](#)). The Scopus database provided information on the year of publication, the authors, the journal and the abstract. We downloaded the full articles as PDF files. Through an in-depth analysis of the entire text and [Supplementary materials](#) of the reviewed research articles, we collected information on the different elements of uncertainty and confidence that will be discussed more in detail in the following section.

Specifically, for all articles, we coded the following information (see also the [Supplementary Excel file](#)):

- the types of uncertainty addressed, and how they were estimated;
- how model performance was estimated; and
- the way confidence was communicated, and the terminology that was used.

The coding was performed by one of the co-authors, reducing the risk for an inhomogeneous treatment of different articles if made separately by several co-authors.

The assumptions and limitations of our study need to be kept in mind while interpreting the results. We searched only one database, used a certain selection of key words and search strings, and our search was made only on articles in English. We reduced the outcome of the search and exclusion process of 193 articles into a random sample of 60 articles. The 60-article sample we analysed represents only a part of the entire SAI literature. To make sure the sample is representative, we checked that the temporal distribution of the sample (i.e. number of articles per publication year) follows that of the original 193 articles (Fig. 1). We also verified that the first authors of the articles in the sample reflect well the diversity of the field (the number of different first authors over the total number of articles was 116/193 in the original sample and 47/60 in the reduced sample). The entire SAI literature should be studied further in future studies and assessment reports with more efficient methods, such as machine learning for text data mining (e.g. [39]).

Three types of modelling uncertainty and how to deal with them

The notion of uncertainty has many definitions, depending on the field of research. Addressing uncertainties is an integral part of research in natural sciences: their sources need to be defined and the uncertainties quantified in the best possible way in order to assess the robustness of the results. In this article, we will refer to uncertainty as 'any departure from the unachievable ideal of complete determinism' [40]. This definition emphasizes the existence of irreducible components of uncertainty in our

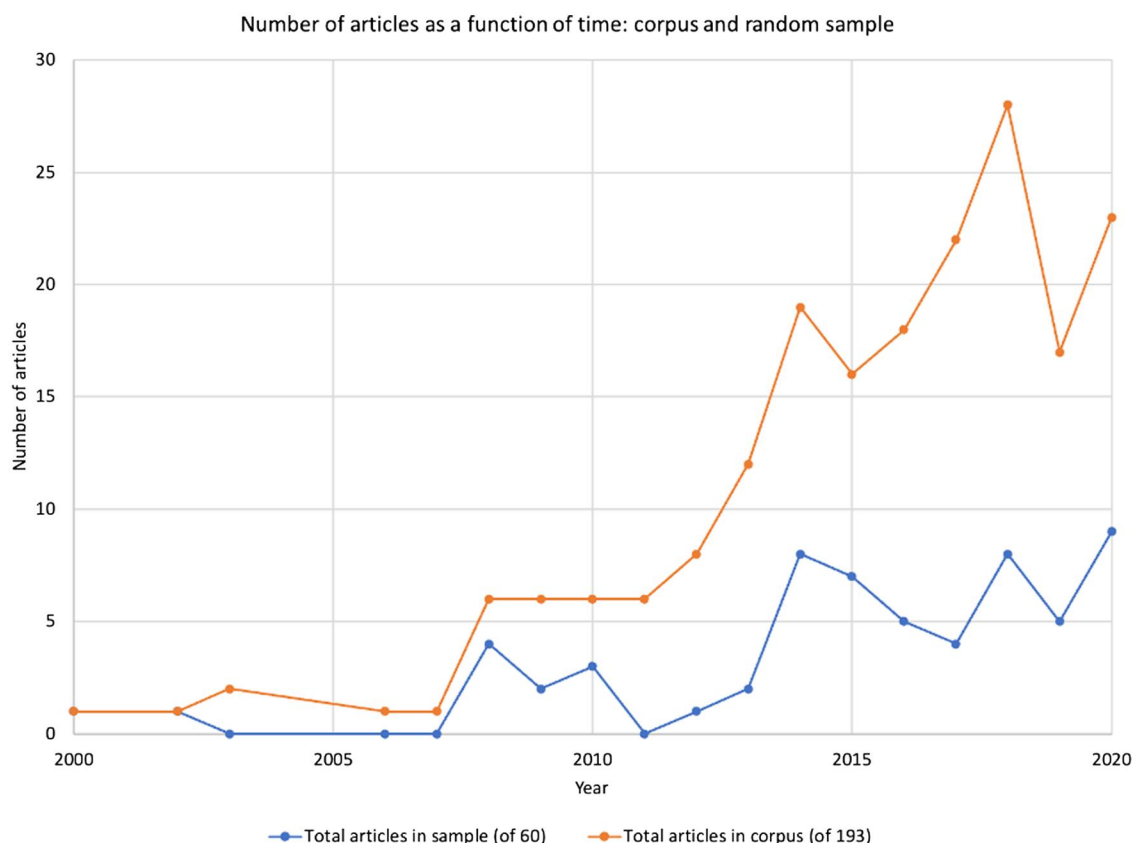


Figure 1. Number of articles in the full corpus of 193 articles (orange, upper curve) and in the sample of 60 articles (blue, lower curve) as a function of publication year.

understanding of the climate system and thus in decision-making about climate change.

Different categorizations are proposed to define sub-classes of the broader concept of uncertainty in the specific context of model-based decision support [40–42]. In this article, we follow the climate modelling community in distinguishing three main sources of uncertainty: model uncertainty, scenario uncertainty, and uncertainty related to internal climate variability [42–45]. These sources of uncertainty are common to all fields using the climate modelling approach, and many of the uncertainties in SRM modelling are the same that arise in modelling future climate change. Moreover, the uncertainties related to aerosol and cloud processes in climate research are very high [1], and as these are the key processes involved in SRM, there is a very strong link between uncertainties in climate and SRM research.

First, model uncertainty arises from the difference between the real climate system and the model used to represent it. There are two main sources of model uncertainty. Our incomplete knowledge about the functioning of the climate system and its processes as well as simplifications made in the representation of the system (e.g. included processes, resolution, parametrization of sub-grid scale processes) produce uncertainty arising from the model structure (model structure uncertainty). The a priori values assigned to model parameters also generate parameter uncertainty [46]. Further uncertainty (unknown unknowns) could arise from processes that are not currently considered but that could play an important role in future climate conditions.

Second, scenario uncertainty, also called driving force uncertainty, reflects the inherent uncertainty on the possible futures described by the used scenarios. A climate scenario is a 'plausible

description of a possible future state of the world' composed of different radiative forcing factors (e.g. forcing corresponding to a particular greenhouse gas emission path, forcing corresponding to the injection of stratospheric aerosols) [47]. When exploring the effects of a certain scenario of solar geoengineering deployment, assumptions are made about (i) the context in which this particular strategy would be deployed (future emission pathways, future radiative forcings, etc.) and (ii) the feasibility of the deployment (technically, politically, socially, economically, etc.). Particularly, the political and social components of feasibility are subjected to a particular type of uncertainties that are difficult to anticipate and quantify. Political and social acceptability is a construct highly dependent on national politics and cultural features. Therefore, a scenario is a single combination of 'what-if' conditions among the infinite range of possible future states of the world [46]. Scenario uncertainty is difficult to quantify, inherently irreducible, and increases the farther we are from present-day conditions.

Third, uncertainty related to internal climate variability reflects that, even in the absence of external forcing, the climate system varies naturally, due to the existence of non-linear dynamic processes. It is therefore also necessary to differentiate internal climate variability and forcing response in climate projections by using long time scales and different initial conditions [48].

There are different ways to quantify each type of uncertainty. First, model uncertainty (model structure and parameter uncertainty) can be explored through different methods. The discussion on and quantification of these uncertainties in sulphate aerosol models, relevant to both climate modelling and SAI started over 25 years ago [49]. The 'Multiple Model Ensemble' (MME) method performs simulations with multiple climate models that

incorporate similar climate processes and use the same set of scenarios and simulate the same time period. This method addresses uncertainty originating from the model structure and parameter values [50]. Using different models with alternative structures and specifications to simulate the functioning of the climate system allows to partially distinguish their consequences on model outputs. These alternative model structures can be seen as different approximations of the real climate system, constrained by complexity and computational cost. However, the extent to which this approach can contribute to the characterization of uncertainty depends on different elements; for instance, interdependence of the chosen models and choices made to build the ensemble [46]. Model ensembles are often opportunistic and rarely designed through a well thought-through approach. The Geoengineering Model Intercomparison Project (GeoMIP) designed standardized scenarios to perform these MME studies on solar geoengineering [51].

Specific sensitivity analysis explores the sensitivity of projections to changes in elements of the model structure or in parameter values. This definition does not include exploration of the sensitivity of projections to changes in scenario parameters. Perturbed Physics Ensemble (PPE) modelling, a subcategory of sensitivity analysis, also sometimes addresses model structure uncertainty, and more specifically, uncertainty regarding parameter values (Parameter uncertainty). PPE uses a single model, and proceeds to multiple model runs differing from one another by perturbations of a limited number of model parameters, under similar scenarios, preferably with large ensemble size and systematic sampling [52]. As the size of the space representing all the combinations of model parameters is tremendous, sampling the full parameter uncertainty in climate models is impossible because of the computational cost and human time required for result analysis. Hence, simpler versions of climate models or emulators are often used in exploring this uncertainty. As PPE are conducted with a single model, the interpretation of results from PPE are constrained by the structural characteristics of the model used [53]. Both approaches, PPE and MME, can be combined to improve the understanding of key uncertainties in modelling. In SAI literature, one study combined and compared MME and PPE to investigate the different types of uncertainties using standardized GeoMIP scenarios and models, and they called for future studies building upon their approach to better understand the uncertainties in modelling of solar geoengineering [54].

Second, scenario analyses address scenario uncertainty. They explore a broad range of possible future conditions [47]. The computational cost and the time needed to analyse the data limit the possibilities to analyse a large number of scenarios. A systematic sampling of the space of possible futures is not possible, but using multiple scenarios allows to explore degrees of freedom and how they might impact the modelled response. Nevertheless, standardizing scenarios, such as those of GeoMIP, or others that have been recently proposed, are also important in order to facilitate the comparison between models and to assess if a certain feature is recurrent in these projections [51, 55, 56]. It should be mentioned that there are SAI studies that specifically focus on scenario uncertainties by exploring multiple SAI scenarios (e.g. [55, 57]).

Third, to further explore uncertainty arising from the internal natural variability of the climate system in the models and try to separate the climate response to forcing from this internal climate variability, several model runs can be carried out with the same model but with different initial conditions. This is called Single-Model Initial Condition (Large) Ensemble Modelling. It allows to quantify the natural climate variability uncertainty with ensemble statistics. The ensemble size is a primordial criterion for the interpretation of the ensemble results as it is

considered to determine their significance. However, there is no unique, ideal ensemble size; the necessary ensemble size depends on many aspects, such as the studied variable and the acceptable level of error [58]. Large ensembles are particularly important for the assessment of regional climate responses, yet the computational cost of this approach limits the extent to which large ensembles are employed [42]. The stratospheric aerosol Geoengineering Large Ensemble Project (GLENS) was created for the purpose of quantifying the natural variability in the modelled climate response to SAI [59]. A more recent effort with a similar approach has been conducted in the ARISE-SAI project [60]. Other, computationally less expensive approaches can be used as well. These generally look at the statistical significance of the change or difference of a mean value of a parameter with respect to internal climate variability.

These different methods make it possible to quantify the different types of uncertainties. They provide a lower bound of the total uncertainty, but computational cost often limits their extensive use.

Model performance and perceived accuracy in climate projections

The characterization of uncertainties and the explicit communication of known limits and assumptions contribute to the perceived accuracy of modelled projections [29]. We distinguish here two dimensions of the model performance: model fit to past and present observations and 'robustness'.

First, model fit is widely used to estimate confidence in the model [61]. This hypothesis consists in extrapolating a model's ability to produce results for past and present-day conditions that fit well the available observational data (model fit) as an indication of its projection skills for multidecadal and century-timescales in the future. Different indicators of model performance (instances of fit) are used to estimate this empirical accuracy and to compare performance between models [62]. A debate exists on whether these different instances of fit should be considered as a confirmation of the entire model or of its adequacy for particular purposes [63]. In this adequacy-for-purpose perspective, climate modellers often justify the choice of a model for a given purpose by its ability to represent particular processes that they consider relevant for the study. This 'empirical accuracy' factor is crucial in the confidence for interpreting climate projections, but it cannot be verified. The inference from model fit has other known limitations, such as the possible existence of compensating biases introduced during the model calibration process to fit the data, and the fact that future climate conditions could be significantly different from those for which the model has been calibrated [61, 64].

Second, the 'robustness' of results from a model projection corresponds to how well these results agree with those obtained with other models or reported in other studies in the literature. In order to be able to state that a result from a model projection is robust thanks to a good agreement with results from other studies, it is necessary to explain why alternative explanations for the agreement are ruled out. For instance, it could be clarified why the agreement between models in a multi-model ensemble is not a result of models sharing common biases due to interdependence between models [46, 65]. In fact, obtaining consistent results from different models does not necessarily mean that they are accurate. The concept of climate model robustness and its different usages have also been discussed in the field of philosophy of science (see e.g. [66]).

In summary, climate model projections are inherently uncertain, as a result of model structure uncertainty, scenario uncertainty, and uncertainty related to internal climate variability. These uncertainties can be characterized and quantified through different methods, helping to establish confidence in the modelled results. Confidence is also related to model performance, expressed as model fit and robustness. Finally, it is also important how uncertainty and confidence are communicated: that is, which terminology is used, and how. All of the above applies to both climate change projections and to other topics, such as SAI, studied with the climate modelling approach.

Results of the systematic literature review

In this section, we provide quantitative and qualitative evidence on the different factors used in the discussion on the confidence and uncertainties in SAI simulations in the studied articles of our sample. We first discuss if and how the different types of uncertainties were studied. Then we address how model confidence was expressed in the studied articles. Finally, we map the vocabulary used for conveying these aspects in the reviewed literature. The full dataset with further details can be found in [Supplementary materials](#).

Types of uncertainties and how they were dealt with

As the sample of analysed articles contained a large variety of non-standardized scenarios and experiments and the articles did not perform quantitative scenario uncertainty estimations, we are not discussing scenario uncertainty in this analysis. Only results concerning model uncertainty and natural climate variability uncertainty are presented in the following, and show that, when uncertainties are discussed, the SAI studies apply a large variety of methods to estimate the different types of uncertainties.

Model structure uncertainty was mainly investigated with Multiple Model Ensemble (MME) using similar scenarios. Among the 60 reviewed articles, 21 performed MME modelling with 2–13 models ([Table 1](#)). Sixteen studies shared common methodologies (e.g. similar set of models and scenarios) as they were part of the Geoengineering Model Intercomparison Project (GeoMIP). In particular, four articles (see [Table 1](#)) used results of eight Earth System Models from GeoMIP [67–70]. Results from MME modelling were communicated and represented in different ways: (i) by focusing on the unweighted multi-model mean with different complementary metrics on the spread of the results, such as the across-model spread (range, under the form of shading, parentheses), the ensemble standard deviation (in the form of an error bar, or \pm) as well as stippling on maps to represent areas where a certain proportion of the involved models agree on the sign of change; and (ii) by representing results of the individual models separately to better identify differences between models.

Ten of the 60 reviewed articles explicitly conducted sensitivity analyses. These sensitivity analyses were implemented by removing (in four articles) entire modules from the climate/impact model structure (e.g. projections with or without interactive

chemistry [71]), or processes (with or without heat and water stress in crop modules [72]), as well by changing (in six articles) parameter values (e.g. sensitivity of crop yields to temperature and precipitation changes, crop failure threshold [73, 74]). Results were often presented through comparing the results of the model runs with varied parameters (e.g. global mean radiative forcing for the different model runs) as curves in a plot or as values tabulated in tables for easy comparison.

Two of the 60 reviewed studies explicitly used a Perturbed Physics Ensemble (PPE) in order to explore the parameter uncertainty. One of them used a crop model with perturbations of 20 parameters [72]. The resulting ensemble grouped 19 parameter sets for the historical climate and 76 for projections (19 parameter sets for 4 different parametrizations of the response to CO₂). The other PPE was conducted with HadCM3, an older generation climate model (CMIP3), with a PPE ensemble size of 20 members, perturbing 8 parameters [54]. The authors explicitly stated that the small ensemble size, the small number of perturbed parameters and the shared model structure undermined the strength of the interpretation of the PPE results [54]. This also applies to the capability of the PPE to inform us on the climate response in a Multiple Model Ensemble [54]. In both PPE articles, methodologies to select tested parameter sets and choose which parameter to perturb were extensively communicated. Results from the PPE using HadCM3, such as the surface air temperature change between geoengineering and control scenarios, were communicated with the ensemble mean complemented with the ensemble standard deviation as well as the range (minimum and maximum of the ensemble) [54]. For graphical representations, the PPE ensemble mean was sometimes shown alone (with or without standard deviation and range) or along with individual PPE members.

Eighteen articles addressed uncertainty related to internal climate variability by carrying out several model runs with the same model but with differing initial conditions. These articles have more than one Single Model Initial Condition Ensemble member with ensemble sizes from 2 to 30 members. Six articles use a Large Ensemble with at least 10 ensemble members. Three of these studies used a large ensemble generated in the context of the Stratospheric Aerosol Geoengineering Large Ensemble (GLENS) Project [9, 75, 76]. Results corresponding to the different ensemble members were represented either (i) by regrouping them with the ensemble mean, complemented by information about the spread (with the range, standard deviation or standard error in shading); or (ii) by communicating the results individually. Other, computationally less expensive methods were used in a majority of publications (42 articles) looking at the statistical significance of the change or difference of a mean value of a parameter providing a significant enough signal with respect to internal climate variability (noise). In some cases, both methods were used (15 articles).

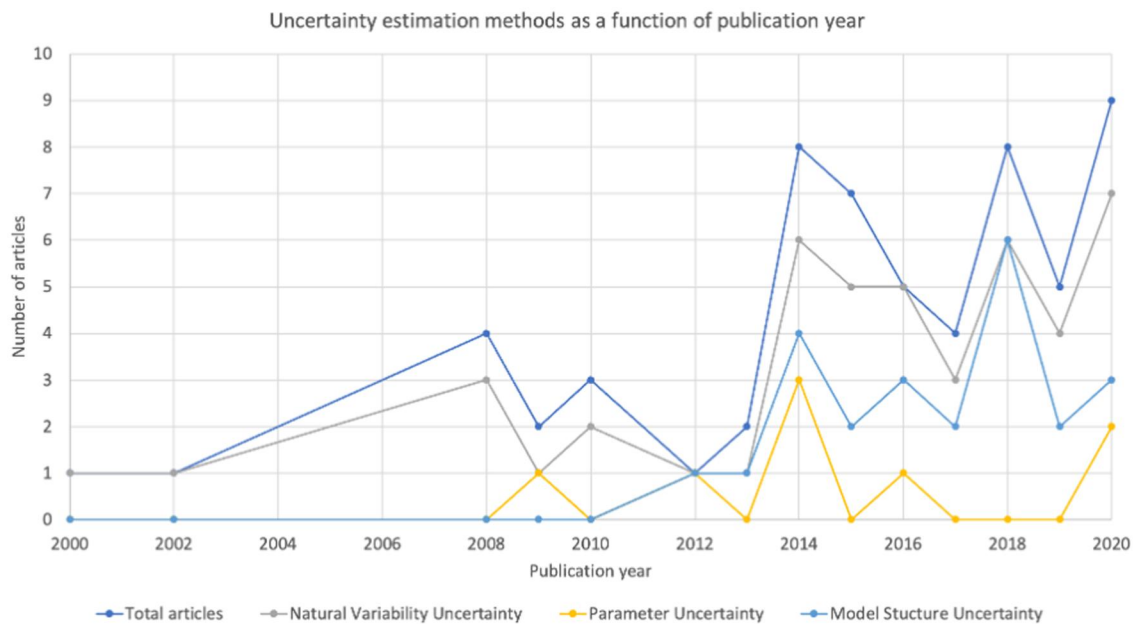
The statistics on the uncertainty estimation methods are given in [Table 2](#). None of the articles systematically addressed all of these different sources of uncertainty. As our sample spans two decades of SAI modelling research, we have also taken a look at the temporal evolution of uncertainty estimation in the articles ([Fig. 2](#)). We note that, in the studied articles, there has been an increase in interest concerning parameter and model structure uncertainty especially since 2010, the earlier studies (before 2010) looking nearly exclusively at natural variability uncertainty. This points to an increasing importance of uncertainty quantification in the field.

Table 1. Number (and percentage) of models used per article in the full sample

Number of models	1	2	4–7	8	9–10	12–13
Number of articles (%)	39 (65.0)	5 (8.3)	4 (6.7)	4 (6.7)	4 (6.7)	4 (6.7)

Table 2. Number/percent of articles using different uncertainty analysis methods

Method	MME	PPE	Sensitivity analysis	Internal variability (ensembles)	Internal variability (signal-to-noise)
Number (%)	10 (16.7)	2 (3.3)	10 (16.7)	18 (30)	42 (70)

**Figure 2.** Evolution of uncertainty estimation methods in the articles as a function of publication year. The total number of articles (dark blue) is plotted in addition to natural variability (grey), parameter (yellow) and model structure uncertainty (light blue).

Estimation of model performance and confidence

In the reviewed publications, indications of model performance were given in three forms. First, articles generally stated that (i) relevant processes were included in the model configuration, and/or (ii) the used model could be considered as being 'state-of-the-art', as it was participating in intercomparison works (CMIP) and assessment reports of the IPCC. Second, 27 articles referred to other publications in which some form of model evaluation in past or present-day conditions had been conducted. Third, 15 articles stated that the study included a comparison of model results to past or present-day observational data or to results from other studies performed with another model of similar or higher complexity. These elements were sometimes communicated with figures (e.g. Fig. 2 in [77]; Fig. 1 in [78]) showing the comparison between a simulation of historical baseline and observational data, sometimes with error bars to show the statistical significance and correlation factor. More often, the indications were expressed with qualitative statements (using terms such as 'good agreement', 'realistic', 'reproduce reasonably well', 'accurately simulate', 'simulate consistently', 'compares very well'). Twenty of the reviewed articles (33% of the sample) did not include any explicit elements about model performance. Furthermore, the reviewed articles mentioning the model performance for past or present-day conditions did not explicitly state that the inference from model fit hypothesis was the basis for the extrapolation to the model's ability to produce reliable projections of the future. Both the lack of explicit elements on model performance and the lack of statements on model fit hypothesis can be explained by the fact that the extent to which models that have been used for a long time in the field are reliable or considered to be reliable for a particular purpose is known to the

authors and to the community as it has been extensively tested and documented in different settings.

Terminology used to communicate uncertainty and confidence

Finally, we also explored how the articles used different notions when speaking about uncertainty and confidence in their results. A variety of terms were employed and we have classified them into three categories: terms related to the methodology of the studies, to the effects of SRM, and to the uncertainty metrics (Fig. 3).

The first category on methodology (red bars in Fig. 3) is also related to the above discussion on model confidence, as terms such as 'assumptions', 'approximation' and 'hypothesis' are found in this category. The results show (blue bars in Fig. 3) that nearly all of the articles refer to effects and impacts of SRM, more than half mention risks, and more detailed terms such as damage and harm were used in 10–20% of the articles. Statements about the effects of SRM could be expected to be accompanied by, for example, an estimation of the probability of such effects taking place or the uncertainties related to the impacts and risks (violet bars). In what follows, we analyse the way the articles use the notions related to uncertainty to back up statements on the effects of SRM.

The adverbs 'likely' and 'unlikely' were used in a qualitative manner in the great majority (54) of the reviewed articles. First, they were used to express authors' confidence in the results of the projections (examples: 'In particular, it is likely that ice loss will be greater than predicted by the icesheet model, resulting in a higher level of SRM geoengineering needed to avert a given sea-level rise' [79]; 'hence the main conclusions reached in this study

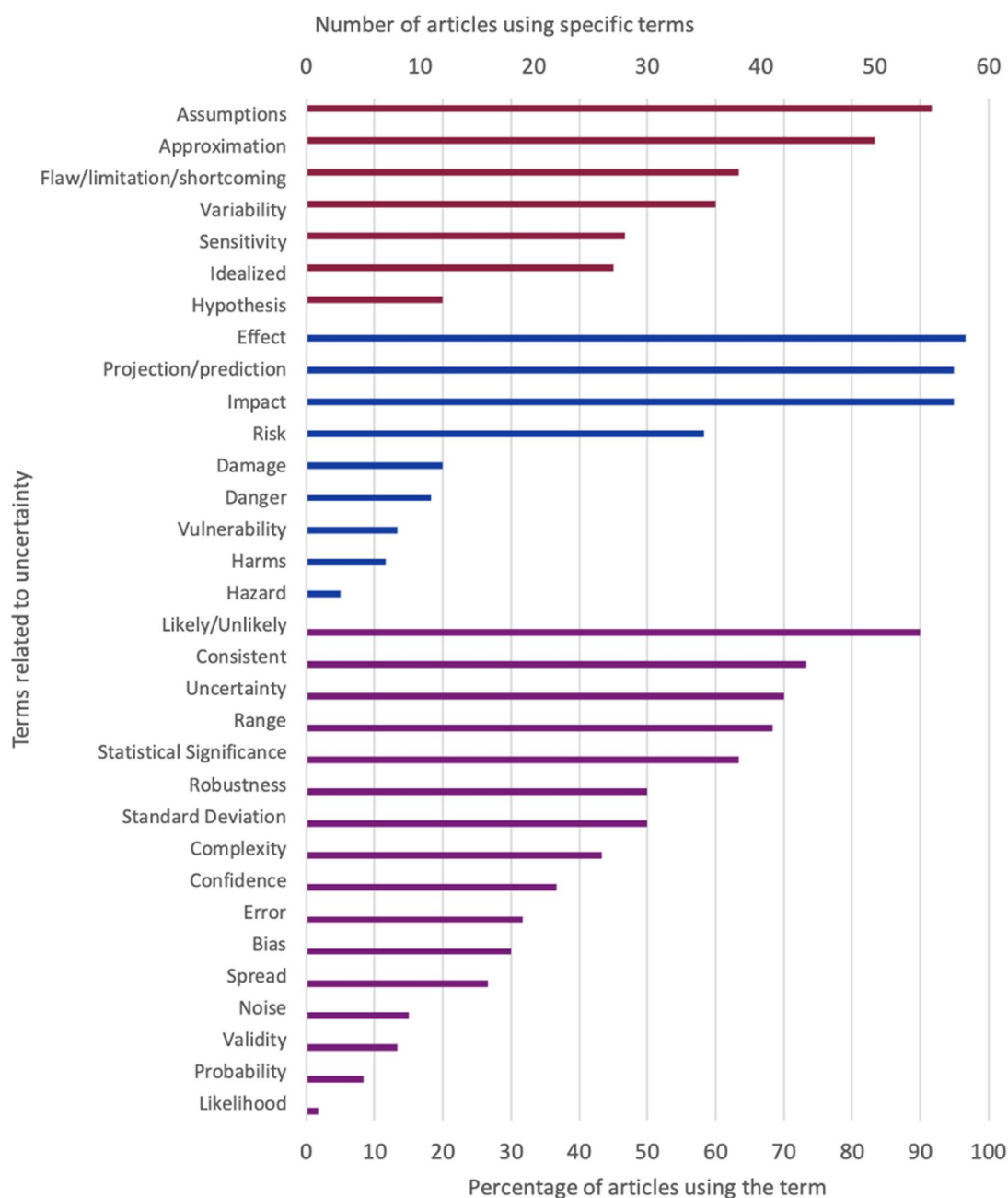


Figure 3. Number (top x-axis) and percentage (bottom x-axis) of articles using the listed terms related to uncertainty. The terms are grouped in three categories: Methodology (red, top group), effects (blue, middle group) and uncertainty metrics (violet, bottom group).

are unlikely to be altered by distributing the aerosols within the stratosphere' [80]). Second, they could also indicate a probable (but not directly demonstrated) explanation for a certain phenomenon showing up in the studied experiments (examples: 'Decreasing plant water use due to a decrease in transpiration is likely driving the increase in global soil moisture' [81]; 'This is likely due to the following reason' [82]).

The term 'Consistent' was used in 44 studies, also with different usages. This term was mainly used (30 articles) to describe how the results agreed with results of other studies in the literature, with (28) or without (2) reference to precise studies. This term was also employed to indicate when there seemed to be an agreement between models on the effects or climate responses that were considered significant (8), sometimes with a defined threshold (e.g. consistency for at least 75% of all models).

Another usage of 'consistent' was to highlight that projected effects and climate responses were coherent between each other in light of expected relationships between them (10) based on our understanding of physics (e.g. 'physically consistent' [75]; 'consistent with Clausius-Clapeyron relation' [83]).

'(Statistical) significance' (38 articles) and 'confidence' (22) were used to describe the statistical significance of the results. Among the 60 reviewed articles, 34 used at least one statistical procedure to test the statistical significance of the results. The procedures included Student's t-test (16), Kolmogorov-Smirnov test (3), Wilcoxon signed-rank test (2), Mann-Whitney U-test (1), Kappa statistics (1), and tests not explicitly described in the article. Metrics communicated in relation to the tests were the p-value interpreted as the 'confidence level' on the statistical significance of the results, for example, $P \leq 0.05$ interpreted as a

confidence level of 95%. Different p -values were used: most frequently 0.05, and rarely 0.01 or 0.1. In most of these studies, only the differences that were deemed statistically significant were further analysed and interpreted.

The notion of ‘robustness’ of results is one of the most frequently used terms, being mentioned in half of the reviewed publications (30 articles), and associated with different criteria. The term is commonly used in climate sciences, especially when describing converging lines of evidence, including but not restricted to convergent predictions of a group of more or less independent models, and has sparked a discussion among philosophers of science. A study has reviewed the philosophical literature on types of robustness, in particular related to climate modelling [66]. They describe a distinct type of ‘model robustness’ that is not only supported by the converging results of the models sharing a common causal core, but also by the independently and empirically verified parameterizations in these models [66]. In practice, this term was used in various ways in the studied articles. ‘Robustness’ was most commonly used in association with results from multi-model ensembles, when there seemed to be a good agreement on a particular climate response between models (15 articles). This was done either without mention of criteria of ‘good’ agreement, or with explicitly mentioned criteria (e.g. 75% of the models [84], or 10 out of 13 models agreeing on the sign of change [85]). This usage of ‘robustness’ was found mainly for results from the GeoMIP project that included up to 13 models, but also in studies using only two models. In contrast, seven studies, particularly those using a single model, mention that further work, with other models, is needed to determine whether the effects produced in their individual study can be considered as ‘robust’. Another usage of ‘robustness’ to describe results on climate response to SAI scenarios is related to the perceived recurrence of projected climate response in other studies in the literature (7 articles), with or without direct references to particular studies. The use of the term ‘robust’ was also found in association with characteristics of the model (e.g. the model resolution, or the model configuration including aerosol microphysics presented as ‘realistic’ [86]), characteristics of the simulation (e.g. long simulations [87]), and of the initial condition ensemble size. For example, a study may have ‘robustly’ separated signal from internal variability (‘noise’), where ‘robust’ is used as synonym to ‘statistically significant’ for differences between projections using control and SRM scenarios (‘This experiment thus depicts a strongly forced case of geoengineering with carbon dioxide concentrations at 1139 ppm, providing a robust signal-to-noise ratio relative to internal variability’ [84]). Seven studies explicitly mention limitations to the ‘robustness’ of their results for different reasons. These included large uncertainties in the representation of particular processes (e.g. El Niño—Southern oscillation [87]; changes in precipitation patterns [88]), current disagreement among studies in the literature, dependence on the climate sensitivity of the model or on the assumptions of the SRM termination [88], the large disagreement between different models used in the study [89], and the small ensemble size of a Perturbed Parameter Ensemble [54]. Six studies used complementary terms (conditional, modal verbs) to moderate the statement about the ‘robustness’ of the results (e.g. ‘is likely to be robust’, ‘relatively robust’, ‘could be robust’, ‘reasonably robust’, ‘If the results and reasoning described above are robust’). The other studies did not use such complementary terms, but included stronger statements (e.g. ‘a robust result is’, ‘is a robust climate response’).

Quantitative ‘likelihoods’ (one article) or ‘probabilities’ (five articles) of the outcomes were rarely produced in these articles,

resulting from the numerous uncertainties and the deterministic nature of the projections [70, 76, 84, 88, 90].

Discussion

As we have seen, the SAI modelling literature deploys a variety of uncertainty estimation methods focusing on the different types of modelling uncertainties. When discussing these uncertainties, we need to keep the nature of the uncertainty sources in mind.

Model uncertainty

Around one third (21 articles) of the reviewed articles were using more than one model in order to explore model uncertainty. However, the multi-model spread is an imperfect measure of uncertainty and should not be overinterpreted, as the spread is a biased estimate of the uncertainty because climate models used in these experiments are generally not independent, sharing more or less important similarities and thus, they can have similar biases (same model structure, same unknowns) [91, 92]. The spread can also result from individual models lacking sufficient descriptions of the relevant processes. Thus, results of individual models are also important to analyse because they can contribute to identifying sources and locations of model uncertainty. This is particularly important in SAI studies where the models represent some of the relevant processes (e.g. aerosol injection) in a simplified way, leading to difficulties in interpreting the inter-model spread in a meaningful way [36, 93, 94]. Also, a sufficient multi-model ensemble size depends on the research question [58]. These ensemble sizes are often limited due to practical reasons, such as computational constraints and available human resources, and as a result of the number of models available to the scientists performing these studies. The ensembles are largely opportunistic in nature rather than constructed in a systematic way to sample model uncertainty, although in relatively large multiple model ensembles, such as GeoMIP, the models include different parametrizations [51]. Thus, the strength of the statements on robustness arising from agreement between models depends on the independence and the diversity of the used models. Furthermore, interpreting equally-weighted multi-model mean as providing more robust projections than individual models relies on the implicit assumption that individual biases of the different models are partially cancelled out in the averaging process. This assumption is an extrapolation of results showing that a multi-model mean generally provides better results for simulation of present-day climate than individual models [46, 95]. It may not be always true: for instance, if weaker models are included in the multi-model ensemble, it could also be that the model spread overestimates the actual uncertainty.

Perturbed Physics Ensemble (2 articles) and Sensitivity Analysis (10 articles) were occasionally used in the reviewed articles to explore uncertainty in model parameters or structure. These methods were only used to explore a limited number of parameters and model features, and did so with older generation models, mainly because of computational constraints. Perturbed physics ensemble may gain back some popularity with the advent of automatic (or machine-assisted) model tuning methods. It has been shown [96] how iterative waves of short simulations could be used to train an emulator to predict a range of model performance metrics. The emulator is thus a key element of this ‘history matching’ procedure to explore the parameter space more systematically and select parameter combinations for which the model emerging properties satisfy some set criteria. Perturbed physics ensemble can then be focused on a small set

of configurations that are equally plausible rather than on a large and arbitrary sample of the parameter space as it is often currently done. This method could be applied to investigate parametric uncertainties on the climate sensitivity [97] but other types of uncertainties (e.g. those related to stratospheric aerosol processes) could be investigated in a similar way.

Natural variability uncertainty

For the investigation of the natural variability uncertainty, initial condition large ensembles remained rare in the sample (only 6 articles with at least 10 ensemble members). Initial condition ensembles require further computing and storage resources and human time for analysing results. These are necessary when there is insufficient signal-to-noise ratio in a single simulation either because the signal is small or the simulation not long enough (e.g. because it is part of a scenario). A key question, which is rarely addressed, is whether such an ensemble is required or not. Furthermore, it is important to remember that internal variability uncertainty also varies from one model to another and is thus impacted by model uncertainty. Any estimates of internal variability uncertainty for a model are therefore biased but nonetheless useful in characterizing different dimensions of uncertainty [98].

Scenario uncertainty

Although scenario uncertainty was not analysed in this study, we note that implications of the different assumptions made in the design of the SAI scenarios were not systematically discussed in the studied literature. However, these scenarios entail assumptions about the technical feasibility (e.g. ability to deliver a projected intervention, no technical failure), human dimensions of uncertainty (e.g. emission trajectories, social/political acceptability of the interventions, global agreement to proceed to an intervention, no disruption of the intervention) as well as other climate forcings such as volcanic eruptions. The reviewed articles using a modelling approach on SAI mainly aimed at gaining a better understanding of the physical aspects of these deployments. However, the assumptions included in the SAI scenarios also create further uncertainty in the interpretation of the projected outcomes. A community approach for producing a range of plausible SAI scenarios to be systematically used in SAI modelling will be helpful for investigating the uncertainties [36, 51, 55, 56].

Model confidence

The terms used for communicating the reliability of the used models and the degree of confidence in the projections are important to choose in a way that adequately reflects the scientific basis of the statements.

Different factors contribute to the communication and thus the perception of the degree of confidence in climate projections, resulting in the use of specific adjectives (e.g. robust, consistent, significant) or adverbs (e.g. likely, unlikely). The meaning of these words and notions used to describe the results of climate projections is of high importance and it is not always clear if they are used in a qualitative or quantitative manner (e.g. using 'likely' with an associated likelihood as is done in the latest IPCC reports, [34]). These words related to the stated degree of confidence can strongly impact the interpretation made of the model projections and thus, the perceived informational value attributed to their results.

A common assumption, not only in the solar geoengineering community but also more generally in climate change research,

is that good model performance for observed past and present-day conditions is an indication of the model's ability to make accurate projections in the future (extrapolation of the model performance). It appears that as the degree of reliability of commonly used and continuously developed models is implicitly known and acknowledged in the field, many (more than a third) of the reviewed articles did not mention the model performance for present-day or past conditions to justify that the model chosen to perform these particular projections was adequate. This is very probably not dissimilar to what is done in modelling studies on climate change. Despite the implicit knowledge in the field, it would be important to explicitly mention and discuss the adequacy of the models for this particular purpose, as the ability of climate models to provide accurate long-term projections cannot be verified and models are calibrated for present-day conditions. Such a discussion could include explanations about (i) the choice of model performance metrics (instances of model fit) used to assess the empirical accuracy of the model for present-day conditions and (ii) the link between these instances of model fit and the assumptions made on the model quality for projecting scenarios in the future [61, 63]. Another element used in justifying the adequacy of a model for this purpose was to state that model configurations included processes that were considered crucial for the quality of the SAI projections (e.g. detailed aerosol chemistry and microphysics). This argument relies on the support from background knowledge about the climate system to strengthen the confidence in the quality of the projections and on the assumption that we have already identified and included all of the processes that could significantly impact results of the climate projections based on SAI scenarios. As a counterexample, most of the current models are missing a detailed treatment of the small-scale physics related to the injection strategy, such as what happens in the plume after the injection, showing that more research and model development is still needed.

Conclusion

Recent assessment reports mention a 'high' level of uncertainty related to SRM, but do not provide systematic or quantitative estimates of the uncertainties [1, 11–13]. The SRM community has called for a more comprehensive investigation of modelling uncertainties (see e.g. [10, 36]). As the field is rapidly evolving there is no doubt that these questions will be investigated. A stocktake of the published literature can provide a baseline for future studies and assessments. Our study provides a quantitative assessment of if and how the uncertainties have typically been estimated in the published literature through a systematic literature review using a representative sample of the SAI modelling literature. Our results show that, overall, the identification and quantification of uncertainties greatly varies between studies, but over time, the SAI literature has started paying increasing attention to uncertainty estimation. We also noticed that assumptions made in the interpretation of results are sometimes considered implicitly known instead of being reported. The reasons for the varying methods of reporting uncertainties and assumptions are probably related to, for example, the computational cost associated with uncertainty quantification, the preliminary and exploratory character of some studies, and implicit knowledge in the community on the adequacy and validity of the models and protocols being used. These elements are important factors to back up statements on the confidence in modelling results and thus in the ability to estimate potential impacts of SAI interventions. Our study is by no means an exhaustive review

and needs to be complemented by others. Moreover, our study does not provide a quantitative assessment of the uncertainties related to SRM: such a full assessment should be accomplished in the future through a wide effort of the whole research community [10, 36].

In the growing community, more studies aim to identify the knowledge gaps. This provides a step towards specifically designed projects that will address the uncertainties in a systematic and comprehensive manner, also ensuring that uncertainties will be more systematically estimated in the literature (see e.g. discussion in [36]). This does not mean that every single article needs to have a comprehensive uncertainty treatment, since some topics are still exploratory (such as the physics that takes place in the injection plume) and their uncertainties can be more comprehensively quantified in later work. In addition, other means of reducing the uncertainties should be kept in mind [10, 23, 24]. Observations of analogous phenomena and their impacts, such as volcanic eruptions or significant changes in tropospheric particle mass (e.g. air pollution measures or reduction of particle emissions by marine transport), can also provide a means to reduce the uncertainties. Better estimates of uncertainties can be acquired with targeted small-scale or laboratory experiments of relevant processes, such as microphysical processes related to aerosol-cloud interactions and the ensuing radiative impact at local scale, or impacts of different materials on atmospheric chemistry. However, such small-scale experiments will not address the question of uncertainties in global scale climate response to solar geoengineering. Moreover, there is no consensus on whether such experiments, specifically targeting solar geoengineering goals, should be performed [99].

Beyond the production of scientific knowledge, in future meta-analyses and assessments, including a discussion on model confidence and establishing a common framework similar to the IPCC guidelines on communication of uncertainties will be of great added value, especially if and when SAI modelling results are to be used in the political arena.

Other types of uncertainties that are somewhat outside the scope of purely climate modelling studies of SRM should not be downplayed. For instance, to avoid any problems related to rapid temperature changes caused sudden discontinuation of SRM ('termination shock', [100, 101]), a resilient SRM infrastructure would be needed for effective global deployment [100]. This could be achieved with a commitment through, for example, international agreements to oversee SRM in the long run and at the planetary scale. Not only scientific but also geopolitical and political uncertainties, such as the emergence of a coalition supporting a non-use agreement, have thus to be evaluated when considering SRM [99, 102, 103]. These issues are also intertwined with the important ethical questions that are raised by SRM [104]. It is of course not expected that all studies address all types of uncertainties. How to combine such uncertainties from a range of studies is thus an important topic for future research and assessments.

Funding

Funding support for this article was provided by the Institut de la transition environnementale de l'Alliance Sorbonne Université (SU-ITE).

Acknowledgements

We are grateful for the constructive comments provided by the anonymous reviewers.

Authors' contributions

Anni Määttänen (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Project administration [lead], Supervision [lead], Visualization [equal], Writing—original draft [equal], Writing—review & editing [lead]), Thibaut Lameille (Conceptualization [lead], Formal analysis [lead], Funding acquisition [supporting], Investigation [equal], Visualization [equal], Writing—original draft [lead]), Carola Kloeck (Conceptualization [equal], Funding acquisition [equal], Methodology [lead], Supervision [equal], Writing—original draft [equal], Writing—review & editing [equal]), Olivier Boucher (Conceptualization [supporting], Writing—review & editing [equal]), François Ravetta (Conceptualization [supporting], Writing—review & editing [equal]).

Supplementary data

Supplementary data is available at Oxford Open Climate Change online.

Data availability

The data underlying this article are available in the article and in its online [supplementary material](#). The article database is provided as a [Supplementary File](#) (Excel format).

References

1. Intergovernmental Panel On Climate Change. *Climate Change 2021—The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. 1st edn. Cambridge, UK and New York, NY, USA: Cambridge University Press, 2023.
2. Reisinger A, Garschagen M, Mach KJ et al. The Concept of Risk in the IPCC Sixth Assessment Report: A Summary of Cross-Working Group Discussions: Guidance for IPCC Authors. Geneva, Switzerland, 2020.
3. IPCC. *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-Industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty*. 1st edn. Cambridge, UK and New York, NY, USA: Cambridge University Press, 2022.
4. Weaver AJ, Zickfeld K, Montenegro A et al. Long term climate implications of 2050 emission reduction targets. *Geophys Res Lett* 2007;**34**:2007GL031018.
5. Crutzen PJ. Albedo enhancement by stratospheric sulfur injections: a contribution to resolve a policy dilemma? *Clim Change* 2006;**77**:211.
6. Rasch PJ, Crutzen PJ, Coleman DB. Exploring the geoengineering of climate using stratospheric sulfate aerosols: The role of particle size. *Geophys Res Lett* 2008;**35**:2007GL032179.
7. Caldeira K, Bala G, Cao L. The science of geoengineering. *Annu Rev Earth Planet Sci* 2013;**41**:231–56.
8. Irvine PJ, Kravitz B, Lawrence MG et al. An overview of the Earth system science of solar geoengineering. *WIREs Clim Change* 2016;**7**:815–33.
9. Irvine PJ, Keith DW. Halving warming with stratospheric aerosol geoengineering moderates policy-relevant climate hazards. *Environ Res Lett* 2020;**15**:044011.
10. Kravitz B, MacMartin DG. Uncertainty and the basis for confidence in solar geoengineering research. *Nat Rev Earth Environ* 2020;**1**:64–75.

11. Committee on Developing a Research Agenda and Research Governance Approaches for Climate Intervention Strategies that Reflect Sunlight to Cool Earth, Board on Atmospheric Sciences and Climate, Committee on Science, Technology, and Law et al. *Reflecting Sunlight: Recommendations for Solar Geoengineering Research and Research Governance*. Washington, D.C.: National Academies Press, 2021, 25762.
12. *Climate Intervention: Reflecting Sunlight to Cool Earth*. Washington, D.C.: National Academies Press, 2015, 18988.
13. United Nations Environment Programme. *One Atmosphere: An Independent Expert Review on Solar Radiation Modification Research and Deployment*. Nairobi, Kenya, 2023.
14. Budyko MI. *Climatic Changes*. Washington, D. C.: American Geophysical Union, 1977.
15. Budyko MI. On present-day climatic changes. *Tellus* 1977; **29**:193–204.
16. Teller E, Wood L, Hyde R. *Global Warming and Ice Ages: I. Prospects for Physics Based Modulation of Global Change*, (UCRL-JC-128715). United States, 1997.
17. Lenton TM, Vaughan NE. The radiative forcing potential of different climate geoengineering options. *Atmos Chem Phys* 2009; **9**:5539–61.
18. Robock A. Volcanic eruptions and climate. *Rev Geophys* 2000; **38**:191–219.
19. Ammann CM, Washington WM, Meehl GA et al. Climate engineering through artificial enhancement of natural forcings: Magnitudes and implied consequences. *J Geophys Res Atmospheres* 2010; **115**:2009JD012878.
20. Robock A, MacMartin DG, Duren R et al. Studying geoengineering with natural and anthropogenic analogs. *Clim Change* 2013; **121**:445–58.
21. Berdahl M, Robock A, Ji D et al. Arctic cryosphere response in the Geoengineering Model Intercomparison Project G3 and G4 scenarios. *J Geophys Res Atmospheres* 2014; **119**:1308–21.
22. Baur S, Nauels A, Nicholls Z et al. The deployment length of solar radiation modification: an interplay of mitigation, net-negative emissions and climate uncertainty. *Earth Syst Dynam* 2023; **14**:367–81.
23. MacMartin DG, Kravitz B. Mission-driven research for stratospheric aerosol geoengineering. *Proc Natl Acad Sci U S A* 2019; **116**:1089–94.
24. Robock A, Bunzl M, Kravitz B et al. A Test for Geoengineering? *Science* 2010; **327**:530–1.
25. MacMynowski DG, Shin H-J, Caldeira K. The frequency response of temperature and precipitation in a climate model. *Geophys Res Lett* 2011; **L16711**, <https://doi.org/10.1029/2011GL048623>.
26. Dykema JA, Keith DW, Anderson JG et al. Stratospheric controlled perturbation experiment: a small-scale experiment to improve understanding of the risks of solar geoengineering. *Phil Trans R Soc A* 2014; **372**:20140059.
27. Doyle EEH, Johnston DM, Smith R et al. Communicating model uncertainty for natural hazards: A qualitative systematic thematic review. *Int J Disaster Risk Reduct* 2019; **33**:449–76.
28. Lemos MC, Kirchhoff CJ, Ramprasad V. Narrowing the climate information usability gap. *Nature Clim Change* 2012; **2**:789–94.
29. Stainforth DA, Allen MR, Tredger ER et al. Confidence, uncertainty and decision-support relevance in climate predictions. *Philos Trans A Math Phys Eng Sci* 2007; **365**:2145–61.
30. McLaren DP. Whose climate and whose ethics? Conceptions of justice in solar geoengineering modelling. *Energy Res Soc Sci* 2018; **44**:209–21.
31. Petr M, Vacchiano G, Thom D et al. Inconsistent recognition of uncertainty in studies of climate change impacts on forests. *Environ Res Lett* 2019; **14**:113003.
32. Kause A, Bruine De Bruin W, Domingos S et al. Communications about uncertainty in scientific climate-related findings: a qualitative systematic review. *Environ Res Lett* 2021; **16**:053005.
33. Kravitz B, Sikka T. Conducting more inclusive solar geoengineering research: A feminist science framework. *Environ Plan E Nat Space* 2023; **6**:1636–60.
34. Mastrandrea MD, Mach KJ, Plattner G-K et al. The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. *Clim Change* 2011; **108**:675–91.
35. De Pryck K. Controversial practices: tracing the proceduralization of the IPCC in time and space. *Glob Policy* 2021; **12**:80–9.
36. Visoni D, Kravitz B, Robock A et al. Opinion: The scientific and community-building roles of the Geoengineering Model Intercomparison Project (GeoMIP) – past, present, and future. *Atmos Chem Phys* 2023; **23**:5149–76.
37. Snyder H. Literature review as a research methodology: An overview and guidelines. *J Bus Res* 2019; **104**:333–9.
38. Xiao Y, Watson M. Guidance on conducting a systematic literature review. *J Plan Educ Res* 2019; **39**:93–112.
39. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Syst Rev* 2019; **8**:163.s13643-019-1074-9.
40. Walker WE, Harremoës P, Rotmans J et al. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integr Assess* 2003; **4**:5–17.
41. Refsgaard JC, Van Der Sluijs JP, Højberg AL et al. Uncertainty in the environmental modelling process—A framework and guidance. *Environ Model Softw* 2007; **22**:1543–56.
42. Lehner F, Deser C, Maher N et al. Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst Dynam* 2020; **11**:491–508.
43. Hawkins E, Sutton R. The potential to narrow uncertainty in projections of regional precipitation change. *Clim Dyn* 2011; **37**:407–18.
44. Yip S, Ferro CAT, Stephenson DB et al. A simple, coherent framework for partitioning uncertainty in climate predictions. *J Climate* 2011; **24**:4634–43.
45. Lovenduski NS, McKinley GA, Fay AR et al. Partitioning uncertainty in ocean carbon uptake projections: Internal variability, emission scenario, and model structure. *Glob Biogeochem Cycles* 2016; **30**:1276–87.
46. Knutti R, Furrer R, Tebaldi C et al. Challenges in Combining Projections from Multiple Climate Models. *J Clim* 2010; **23**:2739–58.
47. Carter TR, La Rovere EL, Jones RN et al. Developing and applying scenarios. In: McCarthy J, Canziani OF, Leary N et al. (eds). *Climate Change 2001: Impacts, Adaptation and Vulnerability*. Cambridge: Cambridge University Press, 2001, 145–90.
48. Kay JE, Deser C, Phillips A et al. The Community Earth System Model (CESM) Large Ensemble Project: a community resource for studying climate change in the presence of internal climate variability. *Bull Am Meteorol Soc* 2015; **96**:1333–49.
49. Pan W, Tatang MA, McRae GJ et al. Uncertainty analysis of direct radiative forcing by anthropogenic sulfate aerosols. *J Geophys Res Atmos* 1997; **102**:21915–24.
50. Sanderson BM. A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. *J Clim* 2011; **24**:1362–77.

51. Kravitz B, Robock A, Tilmes S et al. The Geoengineering Model Intercomparison Project Phase 6 (GeoMIP6): simulation design and preliminary results. *Geosci Model Dev* 2015;**8**:3379–92.
52. Stainforth DA, Aina T, Christensen C et al. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 2005;**433**:403–6.
53. Yokohata T, Annan JD, Collins M et al. Reliability and importance of structural diversity of climate model ensembles. *Clim Dyn* 2013;**41**:2745–63.
54. Irvine PJ, Boucher O, Kravitz B et al. Key factors governing uncertainty in the response to sunshade geoengineering from a comparison of the GeoMIP ensemble and a perturbed parameter ensemble. *J Geophys Res Atmos* 2014;**119**:7946–62.
55. MacMartin DG, Visioni D, Kravitz B et al. Scenarios for modeling solar radiation modification. *Proc Natl Acad Sci U S A* 2022;**119**:e202230119.
56. Lockley A, Xu Y, Tilmes S et al. 18 Politically relevant solar geoengineering scenarios. *SESAMO* 2022;**4**:18127.
57. Tilmes S, MacMartin DG, Lenaerts JTM et al. Reaching 1.5 and 2.0°C global surface temperature targets using stratospheric aerosol geoengineering. *Earth Syst Dynam* 2020;**11**:579–601.
58. Milinski S, Maher N, Olonscheck D. How large does a large ensemble need to be? *Earth Syst Dynam* 2020;**11**:885–901.
59. Tilmes S, Richter JH, Kravitz B et al. CESM1(WACCM) Stratospheric Aerosol Geoengineering Large Ensemble Project. *Bull Am Meteorol Soc* 2018;**99**:2361–71.
60. Richter JH, Visioni D, MacMartin DG et al. Assessing responses and impacts of solar climate intervention on the Earth system with stratospheric aerosol injection (ARISE-SAI): protocol and initial results from the first simulations. *Geosci Model Dev* 2022;**15**:8221–43.
61. Baumberger C, Knutti R, Hirsch Hadorn G. Building confidence in climate model projections: an analysis of inferences from fit. *WIREs Clim Change* 2017;**8**:e454.
62. Gleckler PJ, Taylor KE, Doutriaux C. Performance metrics for climate models. *J Geophys Res Atmos* 2008;**113**:2007JD008972.
63. Parker WS. Confirmation and adequacy-for-purpose in climate modelling. *Aristot Soc Suppl* 2009;**83**:233–49.
64. Reifen C, Toumi R. Climate projections: past performance no guarantee of future skill? *Geophys Res Lett* 2009;**36**:2009GL038082.
65. Sanderson BM, Knutti R, Caldwell P. Addressing interdependency in a multimodel ensemble by interpolation of model properties. *J Clim* 2015;**28**:5150–70.
66. Lloyd EA. Model robustness as a confirmatory virtue: The case of climate science. *Stud Hist Philos Sci Part Sci* 2015;**49**:58–68.
67. Bal PK, Pathak R, Mishra SK et al. Effects of global warming and solar geoengineering on precipitation seasonality. *Environ Res Lett* 2019;**14**:034011.
68. Glienke S, Irvine PJ, Lawrence MG. The impact of geoengineering on vegetation in experiment G1 of the GeoMIP. *J Geophys Res Atmos* 2015;**120**,10,196–10,213, <https://doi.org/10.1002/2015JD024202>.
69. Guo A, Moore JC, Ji D. Tropical atmospheric circulation response to the G1 sunshade geoengineering radiative forcing experiment. *Atmos Chem Phys* 2018;**18**:8689–706.
70. Moore JC, Grinsted A, Guo X et al. Atlantic hurricane surge response to geoengineering. *Proc Natl Acad Sci U S A* 2015;**112**:13794–9.
71. Muthers S, Raible CC, Rozanov E et al. Response of the AMOC to reduced solar radiation—the modulating role of atmospheric chemistry. *Earth Syst Dynam* 2016;**7**:877–92.
72. Yang H, Dobbie S, Ramirez-Villegas J et al. Potential negative consequences of geoengineering on crop production: A study of Indian groundnut. *Geophys Res Lett* 2016;**43**:11786–95.
73. Pongratz J, Lobell DB, Cao L et al. Crop yields in a geoengineered climate. *Nature Clim Change* 2012;**2**:101–5.
74. Yang H, Dobbie S, Ramirez-Villegas J et al. South India projected to be susceptible to high future groundnut failure rates for future climate change and geo-engineered scenarios. *Sci Total Environ* 2020;**747**:141240.
75. Fasullo JT, Tilmes S, Richter JH et al. Persistent polar ocean warming in a strategically geoengineered climate. *Nature Geosci* 2018;**11**:910–4.
76. Odoulami RC, New M, Wolski P et al. Stratospheric aerosol geoengineering could lower future risk of 'Day Zero' level droughts in Cape Town. *Environ Res Lett* 2020;**15**:124007.
77. Jones AC, Haywood JM, Jones A. Climatic impacts of stratospheric geoengineering with sulfate, black carbon and titania injection. *Atmos Chem Phys* 2016;**16**:2843–62.
78. Moore JC, Yue C, Zhao L et al. Greenland ice sheet response to stratospheric aerosol injection geoengineering. *Earths Future* 2019;**7**:1451–63.
79. Irvine PJ, Lunt DJ, Stone EJ et al. The fate of the Greenland Ice Sheet in a geoengineered, high CO₂ world. *Environ Res Lett* 2009;**4**:045109.
80. Modak A, Bala G. Sensitivity of simulated climate to latitudinal distribution of solar insolation reduction in solar radiation management. *Atmos Chem Phys* 2014;**14**:7769–79.
81. Dagon K, Schrag DP. Quantifying the effects of solar geoengineering on vegetation. *Clim Change* 2019;**153**:235–51.
82. Xu Y, Lin L, Tilmes S et al. Climate engineering to mitigate the projected 21st-century terrestrial drying of the Americas: a direct comparison of carbon capture and sulfur injection. *Earth Syst Dynam* 2020;**11**:673–95.
83. Bala G, Duffy PB, Taylor KE. Impact of geoengineering schemes on the global hydrological cycle. *Proc Natl Acad Sci U S A* 2008;**105**:7664–9.
84. Tilmes S, Fasullo J, Lamarque J et al. The hydrological impact of geoengineering in the Geoengineering Model Intercomparison Project (GeoMIP). *J Geophys Res Atmos* **118**(11):036–11,058. <https://doi.org/10.1002/jgrd.50868>.
85. Yu X, Moore JC, Cui X et al. Impacts, effectiveness and regional inequalities of the GeoMIP G1 to G4 solar radiation management scenarios. *Glob Planet Change* 2015;**129**:10–22.
86. Laakso A, Snyder PK, Liess S et al. Differing precipitation response between solar radiation management and carbon dioxide removal due to fast and slow components. *Earth Syst Dynam* 2020;**11**:415–34.
87. Malik A, Nowack PJ, Haigh JD et al. Tropical Pacific climate variability under solar geoengineering: impacts on ENSO extremes. *Atmos Chem Phys* 2020;**20**:15461–85.
88. McCusker KE, Armour KC, Bitz CM et al. Rapid and extensive warming following cessation of solar radiation management. *Environ Res Lett* 2014;**9**:024005.
89. Ji D, Fang S, Curry CL et al. Extreme temperature and precipitation response to solar dimming and stratospheric aerosol geoengineering. *Atmos Chem Phys* 2018;**18**:10133–56.
90. Wei L, Ji D, Miao C et al. Global streamflow and flood response to stratospheric aerosol geoengineering. *Atmos Chem Phys* 2018;**18**:16033–50.
91. Masson D, Knutti R. Climate model genealogy. *Geophys Res Lett* 2011;**38**, L08703, <https://doi.org/10.1029/2011GL046864>.

92. Knutti R, Masson D, Gettelman A. Climate model genealogy: generation CMIP5 and how we got there. *Geophys Res Lett* 2013;**40**:1194–9.
93. Pitari G, Aquila V, Kravitz B et al. Stratospheric ozone response to sulfate geoengineering: Results from the Geoengineering Model Intercomparison Project (GeoMIP). *J Geophys Res Atmos* 2014;**119**:2629–53.
94. Bednarz EM, Visioni D, Kravitz B et al. Climate response to off-equatorial stratospheric sulfur injections in three Earth system models—Part 2: Stratospheric and free-tropospheric response. *Atmos Chem Phys* 2023;**23**:687–709.
95. Reichler T, Kim J. How well do coupled models simulate today's climate? *Bull Amer Meteor Soc* 2008;**89**:303–12.
96. Hourdin F, Williamson D, Rio C et al. Process-based climate model development harnessing machine learning: II. Model calibration from single column to global. *J Adv Model Earth Syst* 2021;**13**:e2020MS002225.
97. Hourdin F, Ferster B, Deshayes J et al. Toward machine-assisted tuning avoiding the underestimation of uncertainty in climate change projections. *Sci Adv* 2023;**9**:eadf2758.
98. Mankin JS, Lehner F, Coats S et al. The value of initial condition large ensembles to robust adaptation decision-making. *Earths Future* 2020;**8**:e2012EF001610.
99. Biermann F, Oomen J, Gupta A et al. Solar geoengineering: The case for an international non-use agreement. *WIREs Clim Change* 2022;**13**:e754.
100. Parker A, Irvine PJ. The risk of termination shock from solar geoengineering. *Earths Future* 2018;**6**:456–67.
101. Rabitz F. Governing the termination problem in solar radiation management. *Environ Polit* 2019;**28**:502–22.
102. Biermann F, Möller I. Rich man's solution? Climate engineering discourses and the marginalization of the Global South. *Int Environ Agreements* 2019;**19**:151–67.
103. Reynolds JL. *The Governance of Solar Geoengineering: Managing Climate Change in the Anthropocene*. 1st edn. New York, NY: Cambridge University Press, 2019.
104. Preston CJ. Ethics and geoengineering: reviewing the moral issues raised by solar radiation management and carbon dioxide removal. *WIREs Clim Change* 2013;**4**:23–37.