Lethargic Response to Aerosol Emissions in Current Climate Models

Citation for the original published paper (version of record):
Storelvmo, T., Heede, U., Leirvik, T. et al (2018). Lethargic Response to Aerosol Emissions in Current Climate Models. Geophysical Research Letters, 45(18): 9814-9823.
http://dx.doi.org/10.1029/2018GL078298

AGU100 ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL078298

Key Points:
- Trends in downward solar radiation measured at ~1,400 surface stations are presented for the last half century
- Historical aerosol emissions support the idea that these observed radiation trends were mainly due to changes in atmospheric aerosol loading
- CMIP5 simulations show negligible solar radiation trends over the same period, raising doubts about their ability to simulate future climate

Supporting Information:
- Supporting Information S1

Correspondence to:
T. Storelvmo,
trude.storelvmo@geo.uio.no

Citation:
Storelvmo, T., Heede, U. K., Leirvik, T., Phillips, P. C. B., Arndt, P., & Wild, M. (2018). Lethargic response to aerosol emissions in current climate models. *Geophysical Research Letters*, *45*, 9814–9823. https://doi.org/10.1029/2018GL078298

Received 12 APR 2018
Accepted 23 AUG 2018
Accepted article online 29 AUG 2018
Published online 19 SEP 2018

Lethargic Response to Aerosol Emissions in Current Climate Models

Trude Storelvmo[1,2], Ulla K. Heede[3], Thomas Leirvik[2], Peter C. B. Phillips[3], Philipp Arndt[4], and Martin Wild[5]

[1]Department of Geoscience, University of Oslo, Oslo, Norway, [2]School of Business, Nord University, Bodø, Norway, [3]Department of Geology and Geophysics, Yale University, New Haven, CT, USA, [4]Department of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden, [5]Institute for Atmospheric and Climate Science, ETH-Zurich, Zurich, Switzerland

Abstract The global temperature trend observed over the last century is largely the result of two opposing effects—cooling from aerosol particles and greenhouse gas warming. While the effect of increasing greenhouse gas concentrations on Earth's radiation budget is well constrained, that due to anthropogenic aerosols is not, partly due to a lack of observations. However, long-term surface measurements of changes in downward solar radiation (SDSR), an often used proxy for aerosol radiative impact, are available worldwide over the last half century. We compare SDSR changes from ~1,400 stations to those from the Coupled Model Intercomparison Project Version 5 global climate simulations over the period 1961–2005. The observed SDSR shows a strong early downward trend followed by a weaker trend reversal, broadly consistent with historical aerosol emissions. However, despite considerable changes to known aerosol emissions over time, the models show negligible SDSR trends, revealing a lethargic response to aerosol emissions and casting doubt on the accuracy of their future climate projections.

Plain Language Summary Observations of incoming solar radiation, as measured at approximately 1,400 surface stations worldwide, show a strong downward trend from the 1960s to the 1980s, followed by a weaker trend reversal thereafter. These trends are thought to be due to changes in the amount of aerosol particles in the atmosphere, and we find support for that here in the temporal evolution of anthropogenic aerosol emissions. This is expected because aerosol particles reflect and/or absorb sunlight back to space and have a net cooling effect on Earth's climate. However, we find that the current generation of climate models simulates negligible solar radiation trends over the last half century, suggesting that they have underestimated the cooling effect that aerosol particles have had on climate in recent decades. Despite this, climate models tend to reproduce surface air temperature over the time period in question reasonably well. This, in turn, suggests that the models are not sensitive enough to increasing greenhouse gas concentrations in the atmosphere, with important implications for their ability to simulate future climate.

1. Introduction

Since preindustrial times, global mean surface temperatures have increased by approximately 1 K. An examination of the global surface temperature record reveals that the rate of warming has varied substantially through time, with periods of accelerated warming punctuated by periods with weak or negligible temperature trend (Hartmann et al., 2013). There is now broad consensus within the climate science community on the root cause of the overall warming, namely, rising levels of atmospheric CO_2, now at 50% above preindustrial levels, driven predominantly by anthropogenic burning of fossil fuels (Cook et al., 2016). Understanding exactly how sensitive Earth's climate is to CO_2 (and other greenhouse gas [GHG]) emissions is critically important for efforts to mitigate future climate change and particularly for efforts to limit warming to less than 1.5 °C, a goal now shared by most countries as stated in the Paris Agreement (United Nations, 2016). Despite recognition of the importance of quantifying Earth's climate sensitivity, that is, the global mean surface temperature increase for a given atmospheric CO_2 increase, this quantity remains elusive. Lack of progress on this issue can be partly attributed to the difficulty of deducing climate sensitivity to CO_2 based on observations. Such efforts have been plagued by the fact that aerosol particles, which have a net cooling effect on climate, have been increasing along with CO_2 and have therefore *masked* some unknown proportion of CO_2-induced warming to date (Andreae et al., 2005; Millar & Friedlingstein, 2018). Two apparent pauses in global warming (in the 1960s and in the 2000s) have both been attributed to aerosol changes (D. M. Smith et al., 2016; Wilcox et al., 2013; Wild, 2016). Representing the cooling effect of aerosol particles in global climate models (GCMs) has proven notoriously challenging, and GCM estimates of aerosol cooling continue to diverge (Boucher et al., 2013).

The overall aerosol effect on climate is often quantified in terms of its *effective radiative forcing* (ERF; Myhre et al., 2013). ERF can be defined in this context as the perturbation to Earth's radiation balance at the top-of-the-atmosphere (TOA) associated with a given change in atmospheric composition — it is negative for anthropogenic aerosols, which cause more solar radiation to be reflected back to space, but positive for CO_2 and other GHGs, which trap more infrared radiation in the Earth system. The underlying mechanisms responsible for the aerosol ERF are not well understood, and there have thus far been few observational constraints on models incorporating these mechanism. One implication of the lack of observational constraints concerns the current generation of state-of-the-art GCMs. These models must be able to broadly capture the surface temperature evolution of the last century in order to have any credibility. Yet they all reasonably do so despite having vastly different reported climate sensitivities (Forster et al., 2013; Kiehl, 2007; Knutti, 2008). Among the GCMs that participated in the Coupled Model Intercomparison Project Version 5 (CMIP5), for instance, the reported equilibrium climate sensitivity (ECS) estimates ranged from 2.0 to 4.5 °C. Clearly, observational constraints that embody aerosol forcing elements are urgently needed to enhance model realism and to help narrow the wide range of current ECS estimates.

For this purpose, we here make use of a frequently used proxy for aerosol forcing, namely, perturbations to the incoming solar radiation at the surface (Cherian et al., 2014). Downward fluxes of solar radiation at the surface (SDSR) have been measured extensively at hundreds of stations worldwide since the midtwentieth century and have been recorded as monthly averages in the Global Energy Balance Archive (GEBA; Wild et al., 2017). Because of atmospheric absorption of solar radiation, aerosol forcing evaluated at the surface differs from that evaluated at the TOA. As such, SDSR changes represent an imperfect proxy for ERF, but the GEBA data set is nevertheless a unique and invaluable data set in this context because of its length and relative consistency. Satellite observations of changes to the net solar radiation at the TOA would be preferable, but reliable long-term records are unfortunately not available. Most GEBA stations do not record separate clear-sky and all-sky SDSR records, so a caveat to the analysis presented here is that cloud changes that are unrelated to aerosol changes could to some extent be responsible for the observed trends. However, previous papers have analyzed SDSR trends from a subset of the GEBA stations that do record separate clear-sky and all-sky data, and they consistently do not find support for the idea that the trends are dominated by cloud changes (Wild, 2012, and references therein). In the following, we will therefore assume that the reported SDSR changes are reliable aerosol ERF proxies. The GEBA data set will be described in more detail in the following section (section 2.1), along with the observational data set used for surface temperature in this study (section 2.2). More details on the CMIP5 model output that we utilized for this study are presented in section 2.3. Thereafter follows a comparison of simulated and observed temperature and SDSR trends globally (section 3.1) and regionally (section 3.2). In section 4 we discuss the implications of our findings and conclude the paper.

2. Methods

2.1. The GEBA Data Set

The GEBA data set consists of monthly mean SDSR values in units of watts per meter squared, as measured by radiometers at more than 2,000 surface stations worldwide (Wild et al., 2017). It contributed to the discovery of the phenomenon now known as *global dimming*, a term that refers to the strong downward trend in SDSR documented in surface measurements worldwide starting in the mid-1950s (Liepert, 2002). Many locations saw a reversal of the dimming trend or a *brightening* in the 1980s (notably Europe and North America), while at other locations the dimming trend continues until the present, for example, in India and China (Wild, 2012). The brightening is mainly attributable to air quality measures implemented in Europe and the United States in the 1960s and 1970s, which caused subsequent sharp declines in aerosol emissions in those regions (Monks et al., 2009), leading to an inverse U-shaped effect that is known in economics as the environmental Kuznets curve (Grossman & Krueger, 1995).

The GEBA data quality is ensured through rigorous quality control, as described in Gilgen and Ohmura (1999) and Gilgen et al. (1998). Irrespective of data quality, comparisons between data from a single/few station(s)

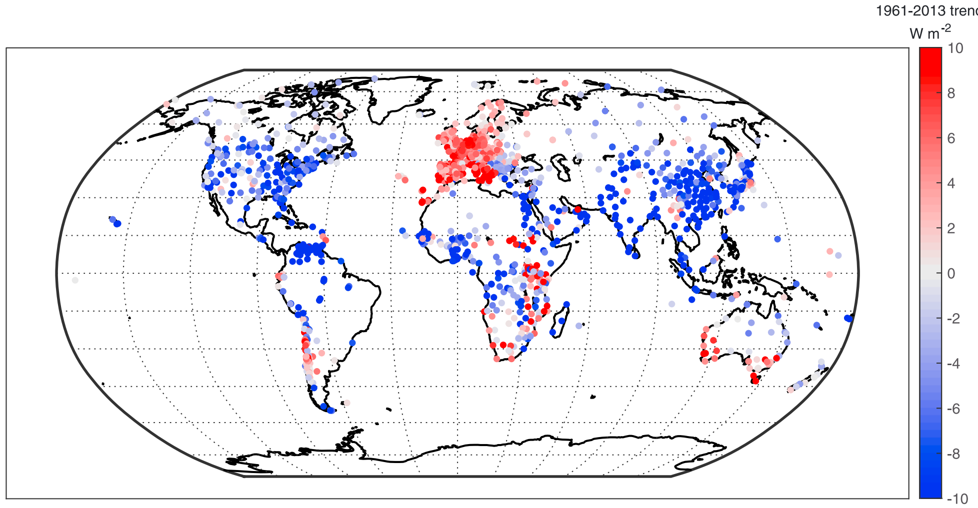

Figure 1. Global overview of location of Global Energy Balance Archive stations included in this study, along with the sign and magnitude of their downward flux of solar radiation at the surface trend, calculated based on 5-year running means from 1961 to 2013 (red corresponds to brightening and blue to dimming). Station records have been complemented with machine-learning algorithms (*random forests*).

and grid box averages are always challenging. The question of whether GEBA stations are representative of their general surroundings, and specifically whether urban impacts dominate SDSR trends, has previously been addressed by Wang et al. (2014), who concluded that the urban impact is very small. This is also consistent with our own finding that separating the GEBA stations into *urban* and *rural* categories did not yield significantly different SDSR trends (see supporting information Figure S3 for further information on the station classification). We therefore deem the comparison between data from the GEBA stations and simulated averages from model grid boxes in which the stations are located as appropriate.

It is worth noting that many of the GEBA stations do not have unbroken time series over the time period studied here. In fact, data availability prevented us from incorporating measurements prior to 1961 in this study, which is based on approximately 1,400 surface stations and has a global as well as regional focus. For the analysis presented here, we produced annual SDSR means for each station based on the monthly mean data for the time period 1961 to 2015. Figure 1 shows the geographical distribution of the stations along with the sign of their individual SDSR trend over the time period considered. It is reassuring that within a given region, the sign of the trend is consistent across stations. The fact that observed trends in pan evaporation, sunshine duration, and diurnal temperature range are generally consistent with the SDSR trends represents a further vote of confidence for the GEBA data (Wild et al., 2005). Whenever stations had measurement gaps, interpolation with the use of a machine-learning algorithm (*Random Forests*; Breiman, 2001) was applied, utilizing the full suite of variables in the Climate Research Unit (CRU) TS4.00 data set described in the following subsection. An alternative to filling data gaps in this way would be to simply carry out the analysis with the gaps present in both the observed and simulated data sets. This approach yields a qualitatively similar outcome to the one presented here; see Storelvmo et al. (2016) for the observed SDSR temporal evolution in this case.

2.2. The CRU TS4.00 Data Set

The CRU high-resolution-gridded climate data set TS4.00 was obtained by interpolating surface observations from meteorological stations across the world's land areas into 0.5° latitude/longitude grid cells (Harris et al., 2014). The CRU TS4.00 data set contains multiple meteorological variables, but for the purpose of this study we use only monthly mean surface air temperature (TS). For each monthly mean SDSR at a given GEBA station, a corresponding TS value from the CRU TS4.00 grid cell within which the station lies is assigned. As for the GEBA data, we create annual mean TS data and analyze the period 1961–2015.

2.3. The CMIP5 Historical Simulations

For comparison with the GEBA and CRU TS4.00 data described above, we use output from the CMIP5 historical experiment (Taylor et al., 2012). The historical simulations are performed either with fully coupled atmosphere-ocean GCMs (AOGCMs) or Earth System Models (ESMs) and are initiated from the respective

Table 1
Coupled Model Intercomparison Project Version 5 Models, Number of Ensemble Members, and Station-Averaged TS and SDSR Linear Trends

Model	# of ensemble members	TS trend (K)	SDSR trend (Wm^{-2})
ACCESS1.0	3	0.90	0.20
ACCESS1.3	3	0.54	1.60
BCC-CSM1.1	3	0.03	−0.08
BNU-ESM	1	1.25	2.14
CanESM2	5	1.15	1.15
CCSM4	3	1.07	0.77
CESM(CAM5)	3	0.79	0.83
GFDL-CM3	5	0.78	−0.28
CMCC-CM	1	0.99	0.03
CMCC-CESM	1	0.60	−0.62
CMCC-CMS	1	1.09	1.07
CNRM-CM4	8	0.85	−0.05
CSIRO-Mk3.6.0	9	0.72	0.35
GFDL-ESM2G	1	1.04	1.26
GFDL-ESM2M	1	0.76	0.65
FGOALS-g2	5	−0.20	0.27
FIO-ESM	3	0.56	−0.14
GISS-E2-H	4	0.64	−2.11
HadCM3	10	0.67	0.21
HadGEM2-CC	1	0.39	−0.48
HadGEM2-ES	4	0.82	0.99
INM-CM4	1	0.63	0.40
IPSL-CM5A-LR	4	1.22	1.52
IPSL-CM5B-LR	1	0.92	0.98
IPSL-CM5A-MR	3	1.03	1.07
MIROC5	5	−0.01	−0.03
MIROC5-ESM	3	0.66	2.17
MPI-ESM-LR	3	0.88	0.68
MPI-ESM-MR	3	1.02	0.63
MPI-ESM-P	2	1.08	0.56
MRI-GCM3	5	0.48	−0.19
NorESM1-M	3	0.71	0.56
Model ensemble mean	108 (sum)	0.75	0.50
Observations	—	0.96	−2.56

Note. SDSR = downward flux of solar radiation at the surface; TS = surface air temperature.

preindustrial control simulations for each AOGCM/ESM. The simulations begin in 1850 and end in 2005 and are forced with observed atmospheric composition and land cover changes. Some models have been run only once for the period 1850–2005, while others have been run repeated times in order to produce an ensemble of simulations. The ensemble members for a given model differ only in their initial states, which are taken from different points in the preindustrial control simulation. In this study, we analyzed monthly mean data from 33 unique AOGCMs/ESMs, which each have one to nine ensemble members, for a total of 108 simulations. The models, their number of ensemble members, and their station-mean temperature and SDSR trends are listed in Table 1. In order to make a fair comparison with the observations, we only utilize output from model grid boxes that contain at least one GEBA station.

10.1029/2018GL078298

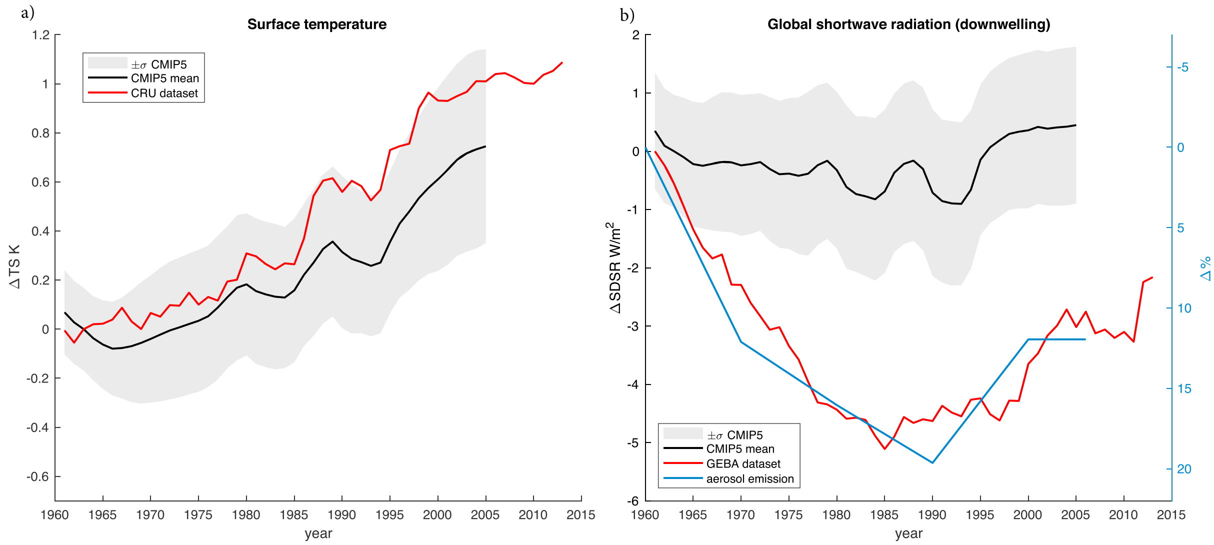

Figure 2. The red line represents the 5-year running mean of changes to temperature (a) and radiation (b) since 1961 averaged across all GEBA stations (b) and corresponding grid boxes for the CRU data set (a). The black line represents the 5-year running mean of the averaged change in temperature (a) and radiation (b) simulated by all 108 CMIP5-model ensembles since 1961 for grid boxes corresponding to the location of GEBA stations. The shaded area represents the running mean of the standard deviation of the trend among models. The mean of the first 5 years of the time series (1961–1966) is calculated for each model and is subsequently subtracted from each year, such that each model has a trend only relative to its individual baseline. The standard deviation thus represents the deviation of a modeled quantity relative to individual model baselines. The right axis and blue line in (b) illustrates linearly interpolated decadal relative changes to total anthropogenic aerosol emission since 1960 averaged over all grid boxes corresponding to the location of GEBA stations based on the Intergovernmental Panel on Climate Change Fifth Assessment Report emission data set (Lamarque et al., 2010). CMIP5 = Coupled Model Intercomparison Project Version 5; CRU = Climate Research Unit; TS = surface air temperature; SDSR = downward flux of solar radiation at the surface; GEBA = Global Energy Balance Archive.

3. Results

3.1. Global Trends in Models and Observations

3.1.1. Temperature Trends

Figure 2a shows the year-to-year change in TS in observations as well as in the ensemble mean of all 108 simulations for the period 1964–2005, both averaged across the ~1,400 stations. Despite considerable spread among the simulations, the ensemble mean TS trend captures the observed trend reasonably well until the early 1980s. Thereafter, the models underestimate the station mean warming. This finding appears to contradict earlier comparisons of simulated and observed global mean surface temperature trends, for which the CMIP5 models actually tend to produce slightly too much warming (Flato et al., 2013). However, the comparison here is somewhat unusual in the sense that we compare year-to-year changes averaged across GEBA stations, which are unevenly distributed around the globe (see Figure 1). The present comparison will therefore likely accentuate systematic regional/continental model biases that may be averaged out in global means. A similar underestimate of warming over land surfaces has previously been reported for CMIP3 climate models (Wild, 2009). In section 3.2 we discuss how the discrepancy between simulated and observed temperature trends in recent decades differ between regions and propose plausible causes for it in section 4.

3.1.2. Surface Radiation Trends

Figure 2b shows the corresponding SDSR changes for the same time period and reveals an egregious mismatch between models and observations. While the observed SDSR decreases steadily from 1961 to ~1990, followed by a weaker upward trend from 1990 to 2005 and beyond, the models show very weak trends that are not statistically significant. These findings are consistent with what has previously been reported for isolated regions, for example, for China and Japan by Allen et al. (2013) and Dwyer et al. (2010) and for Europe by Allen et al. (2013). As evident from Figure 2b, not only does this discrepancy persist in station-averaged SDSR but also its magnitude is striking. This raises major concerns about the fidelity of the models and their representation of processes that govern SDSR. Along with the SDSR changes with time, the figure also shows how the total anthropogenic emissions of aerosol particles and their precursors have changed over the time period of interest (Lamarque et al., 2010). Though we cannot prove a causal relationship, Figure 2b strongly suggests that the SDSR changes were driven by aerosol-induced changes in the reflection and/or absorption of sunlight. In other words, the observed changes in SDSR are broadly consistent with known emission

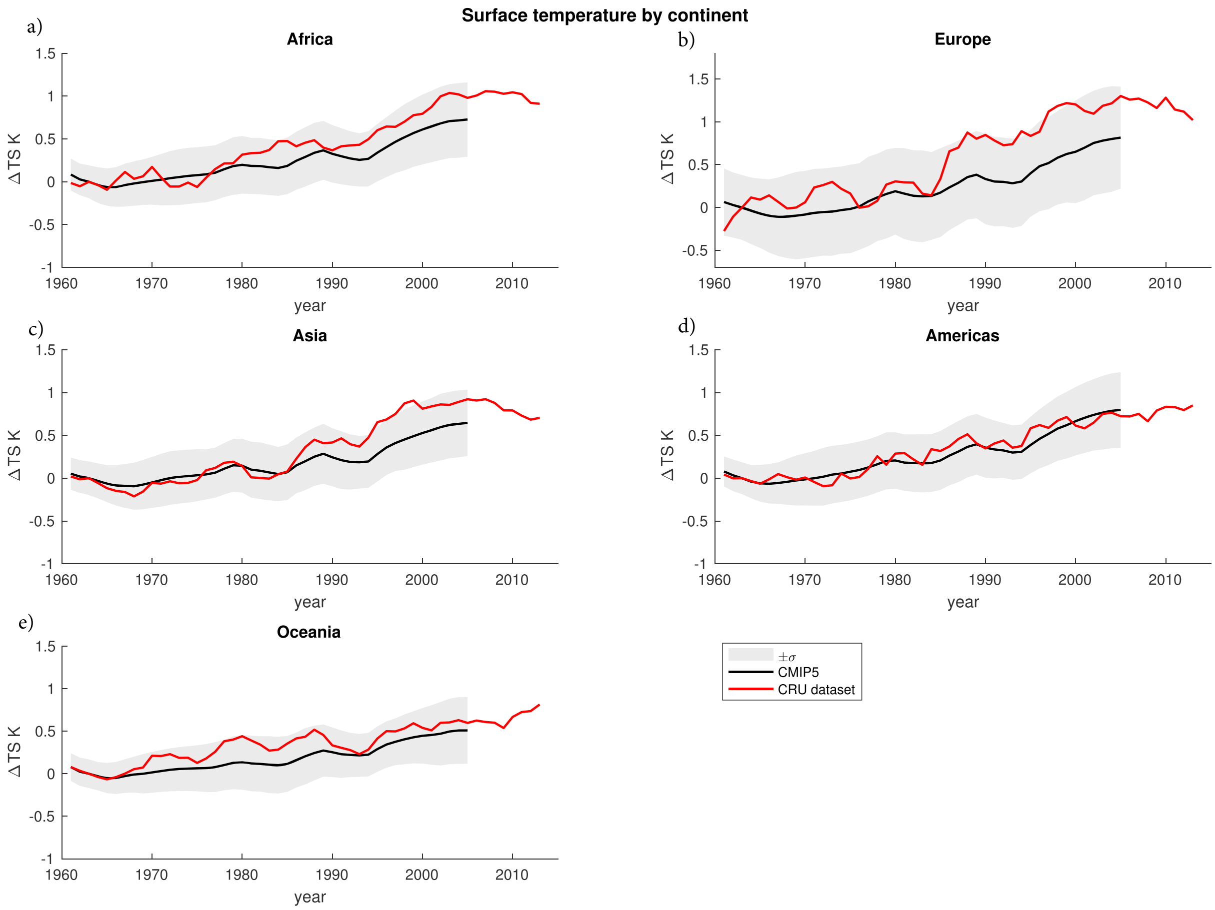

Figure 3. Following the same principles as Figure 2a, the observed and simulated temperature trends for grid boxes corresponding to the Global Energy Balance Archive stations are divided by continent. The two American continents are combined due to the low density of stations in South America (see Figure 1). Stations on oceanic islands far from continents are excluded. TS = surface air temperature; CMIP5 = Coupled Model Intercomparison Project Version 5; CRU = Climate Research Unit.

changes, while the simulated SDSR changes are not. This is true whether SDSR changes are calculated for all-sky or clear-sky conditions (see Figure S1). The lack of simulated SDSR trend is particularly striking given that the emissions displayed in Figure 2b are exactly the same as those used as input to the CMIP5 simulations. As one would expect, selecting only models that have a reported strong negative aerosol ERF (Boucher et al., 2013) slightly improves the comparison with observations (see Figure S2), consistent with findings by, for example, Baker et al. (2015). The following subsection presents the same analysis by region and thus sheds further light on the underlying cause of the disagreement between observed and simulated SDSR trends.

3.2. Regional Trends in Models and Observations

A regional comparison of simulated and observed temperature changes (Figure 3, station averages) reveals that the underestimated warming trend in recent decades stems mainly from Europe, Asia, and Africa. The rapid warming in Europe since the early 1980s is naturally partly attributable to increasing GHG concentrations, but evident in Figure 4a is a coincident positive trend in SDSR across the region, which is not fully captured by the CMIP5 models. Europe is by far the region with the best spatial station coverage (Figure 1), and the reported SDSR trends have been corroborated with satellite observations (Wild et al., 2017). It has been inferred that this observed radiation trend contributed roughly as much warming in recent decades as GHGs (Philipona et al., 2009). This brightening is consistent with known trends in European aerosol emissions and is likely in part responsible for the accelerated warming (Cherian et al., 2014).

Asian emissions are thought to have plateaued in the mid-1990s (Figure 4c), at which point the observed warming trend overshoots the simulated one (Figure 3c). Since the simulated SDSR evolution appears to be completely nonresponsive to the aerosol emissions that are driving the simulations, there is presumably also

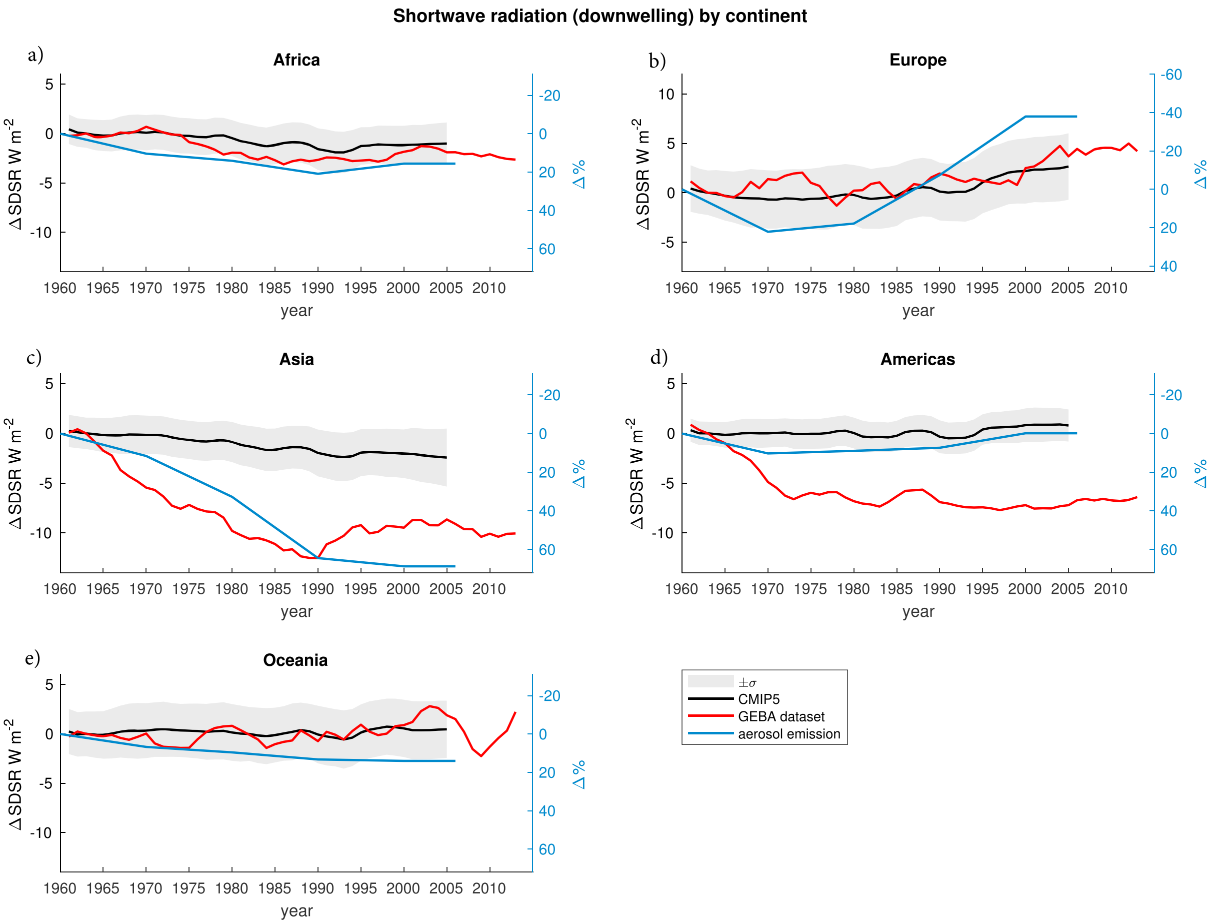

Figure 4. Following the same principles as Figure 2b, the observed and simulated radiation trends for GEBA stations are divided into continents. The two American continents are combined due to the low density of stations in South America (see Figure 1). The blue line corresponds to the right axis and represents the average change relative to the 1960s to anthropogenic aerosol emission over each continent. Note that areas with low density of GEBA stations are excluded such that the aerosol emission changes for each continent is centered around areas of high density of stations, for example, aerosol emissions over Northern Asia, Eastern Sahara, and Eastern South America are not included. SDSR = downward flux of solar radiation at the surface; CMIP5 = Coupled Model Intercomparison Project Version 5; GEBA = Global Energy Balance Archive.

minimal aerosol impact on the simulated temperature evolution. Despite this lack of aerosol cooling, Asian temperature changes are reasonably reproduced by the CMIP5 models until the mid-1990s, implying that they are not sensitive enough to GHG changes. As one would expect if this is the correct explanation, the simulated and observed temperatures start to diverge once Asian aerosol cooling stabilizes in the 1990s. It is worth noting that post-2005 the observations again suggest additional dimming and reduced warming, which is consistent with the current understanding of historical emissions in the region (Hoesly et al., 2018). A recent comprehensive evaluation of Chinese SDSR trends reported by the GEBA stations confirmed that the strong downward trend seen until the 1990s can be trusted but that the trend reversal may be slightly exaggerated (Yang et al., 2018). Identifying the reason(s) that the discrepancy between observed and simulated SDSR trends is particularly large for Asia will require extensive follow-up research, but the fact that several studies have recently concluded that present-day Asian aerosol emissions are underestimated by at least 50% can likely explain some of the discrepancy (e.g., Jiang et al., 2013). This is broadly consistent with reports that a majority of GCMs underestimate aerosol optical depth over East Asia (Shindell et al., 2013), and indeed, historical aerosol emissions for the region have been revised upward for recent decades in the inventories that will be used in the upcoming CMIP phase 6 (Hoesly et al., 2018). However, preliminary results (not shown) suggest that emission increases far exceeding 50% would be required to resolve the disagreement between simulations and observations in this case. In other words, it will likely be necessary to reexamine the GCMs' treatment of processes that govern aerosol lifetime, transport, size distribution, and composition in order to fully resolve this issue.

For Africa and Oceania, SDSR trends are much weaker, as expected given the relatively weak historical emission changes in those regions (but note that according to Wild et al., 2005, the Australian stations have issues that preclude reliable trend calculations there). The observations show moderate *dimming* in Africa, which is practically absent in the CMIP5 simulations. Despite this, the models show slightly less warming than observed, again supporting the notion that they are not sufficiently sensitive to GHG changes. The temperature evolution for the Americas is reasonably well reproduced by the CMIP5 models, even though they appear to lack the observed SDSR trend. However, in this case the observed strong negative SDSR trend cannot fully be explained by the aerosol emissions of Lamarque et al. (2010). That being said, the sharp dimming seen in the GEBA data until the early 1970s, as well as the subsequent stabilization, is supported by the temporal evolution of anthropogenic sulfur dioxide emissions reported by S. J. Smith et al. (2011). We speculate that sulfate was more central to SDSR trends observed over the Americas than they were elsewhere, where the total emissions of aerosols and aerosol precursors are perfectly consistent with observed SDSR trends. In general, all regional SDSR trends could also be affected by any changes in interregional aerosol transport associated with circulation changes on decadal time scales.

4. Discussion and Conclusions

We report a like-for-like comparison of 1961–2005 simulated (CMIP5) air temperature and incoming solar radiation (SDSR) at the surface with those reported for ~1,400 stations by the CRU TS4.00 and GEBA data sets, respectively. The observed temperature evolution is reasonably reproduced by the CMIP5 models until the 1980s, while the overall warming trend is underestimated thereafter. In strong contrast to the observations, the CMIP5 models show little or no trend in SDSR over the relevant time period. This strongly suggests that the models are collectively not responding to aerosol emissions in an appropriate way. The observed SDSR trends are qualitatively consistent with known emission trends for aerosol particles and their precursors, while the simulated trends are not. We therefore argue that the explanation for the SDSR trend discrepancy is not predominantly emission biases but rather errors in the treatment of processes that translate aerosol emissions into clear-sky and all-sky radiative forcings in the models. The failure by the CMIP5 models to reproduce the observed SDSR trend can to some extent explain their biases in simulated temperature trends, especially if a general model underestimation of the temperature response to GHG changes is invoked. This is the most plausible, but not the only possible, explanation for our findings. Other possibilities include the following: (1) If the observed SDSR trends are mainly caused by black carbon particles, which are believed to have a net warming effect, that would mean that dimming does not necessarily equate to cooling; in which case the above logic fails. However, even the largest historical black carbon emission estimates to date do not support this idea (Bond et al., 2013). (2) Even if we are correct to assume that the CMIP5 models underestimate aerosol cooling over land, they could still hypothetically have the correct amount of aerosol cooling (and therefore a realistic ECS) in a global average sense, if they simultaneously overestimate aerosol cooling over the ocean (e.g., due to unrealistic aerosol transport), and (3) the observed SDSR changes could in theory be driven by changes in cloudiness and not aerosols. Cloud feedbacks, in which clouds change in response to climate change and thus further amplify/dampen these changes (Boucher et al., 2013), could certainly be a contributing factor in long-term SDSR changes, but it is very difficult to see how the SDSR trend reversals seen in the observational record could have been caused by cloud feedbacks because all regions considered show persistent warming throughout the time period of interest.

Comprehensive testing of these alternative hypotheses will require extensive follow-up research, which should include further observational analysis as well as climate model sensitivity experiments, the latter ideally conducted in the form of multimodel intercomparison initiatives. This work, some of which is ongoing, goes beyond the scope of the present study. We close here by concluding that the analysis presented supports the idea that CMIP5 models do not adequately represent the climate impact of aerosol changes and that their biased surface temperature trends are thus most easily explained by a lack of aerosol cooling, which is in turn imperfectly compensated through a general underestimation of climate sensitivity.

Acknowledgments
This work was partially supported by a grant from the European Research Council (ERC grant 758005) and a grant from the Research Council of Norway (grant 281071). The data sets analyzed in this study are available for download at the following websites: GEBA (http://www.geba.ethz.ch/) and CRU TS4.00 (https://crudata.uea.ac.uk/cru/data/hrg/).

References

Allen, R. J., Norris, J. R., & Wild, M. (2013). Evaluation of multidecadal variability in CMIP5 surface solar radiation and inferred underestimation of aerosol direct effects over Europe, China, Japan, and India. *Journal of Geophysical Research: Atmosphere*, *12*, 6311–6336. https://doi.org/10.1002/jgrd.50426

Andreae, M. O., Jones, C. D., & Cox, P. M. (2005). Strong present-day aerosol cooling implies a hot future. *Nature*, *7046*, 1187–1190.

Baker, L. H., Collins, W. J., Olivi, D. J. L., Cherian, R., Hodnebrog, Ø., Myhre, G., & Quaas, J. (2015). Climate responses to anthropogenic emissions of short-lived climate pollutants. *Atmospheric Chemistry and Physics*, *15*, 8201–8216.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., et al. (2013). Bounding the role of black carbon in the climate system: A scientific assessment. *Journal of Geophysical Research: Atmospheres*, *118*, 5380–5552. https://doi.org/10.1002/jgrd.50171

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., et al. (2013). Clouds and aerosols: Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 571–657). Cambridge, UK and New York: Cambridge University Press.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Cherian, R., Quaas, J., Salzmann, M., & Wild, M. (2014). Pollution trends over Europe constrain global aerosol forcing as simulated by climate models. *Geophysical Research Letters*, *41*, 2176–2181. https://doi.org/10.1002/2013GL058715

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., et al. (2016). Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environmental Research Letters*, *11*, 48002.

Dwyer, J. G., Norris, J. R., & Ruckstuhl, C. (2010). Do climate models reproduce observed solar dimming and brightening over China and Japan? *Journal of Geophysical Research*, *115*, D00K08. https://doi.org/10.1029/2009JD012945

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W. J., et al. (2013). Evaluation of climate models. In *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*: Cambridge University Press (pp. 741–882). Cambridge, UK and New York: Cambridge University Press.

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L. S., & Zelinka, M. (2013). Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *Journal of Geophysical Research: Atmospheres*, *118*, 1139–1150. https://doi.org/doi:10.1002/jgrd.50174

Gilgen, H., & Ohmura, A. (1999). The Global Energy Balance Archive. *Bulletin of the American Meteorological Association*, *80*, 831–850.

Gilgen, H., Wild, M., & Ohmura, A. (1998). Means and trends of shortwave irradiance at the surface estimated from Global Energy Balance Archive data. *Journal of Climate*, *11*, 2042–2061.

Grossman, G. M., & Krueger, A. B. (1995). Economic growth and the environment. *The quarterly journal of economics*, *2*, 353–377.

Harris, I. P. D. J., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations—The CRU TS3. 10 dataset. *International Journal of Climatology*, *3*, 623–642.

Hartmann, D. L., Tank, A. M. G. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y. A. R., et al. (2013). Observations: Atmosphere and surface: Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., et al. (2018). Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geoscientific Model Development*, *11*, 369–408.

Jiang, H., Liao, H., Pye, H. O. T., Wu, S., Mickley, L. J., Seinfeld, J. H., & Zhang, X. Y. (2013). Projected effect of 2000-2050 changes in climate and emissions on aerosol levels in China and associated transboundary transport. *Atmospheric Chemistry and Physics*, *13*, 7937–7960.

Kiehl, J. T. (2007). Twentieth century climate model response and climate sensitivity. *Geophysical Research Letters*, *34*, L22710. https://doi.org/10.1029/2007GL031383

Knutti, R. (2008). Why are climate models reproducing the observed global surface warming so well? *Geophysical Research Letters*, *35*, L18704. https://doi.org/10.1029/2008GL034932

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., et al. (2010). Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application. *Atmospheric Chemistry and Physics*, *15*, 7017–7039.

Liepert, B. G. (2002). Observed reductions of surface solar radiation at sites in the United States and worldwide from 1961 to 1990. *Geophysical Research Letters*, *29*(10), 1421. https://doi.org/10.1029/2002GL014910

Millar, R. J., & Friedlingstein, P. (2018). The utility of the historical record for assessing the transient climate response to cumulative emissions. *Philosophical Transactions of the Royal Society A*, *376*, 20160449.

Monks, P. S., Granier, C., Fuzzi, S., Stohl, A., Williams, M. L., Akimoto, H., et al. (2009). Atmospheric composition change—Global and regional air quality. *Atmospheric Environment*, *33*, 5268–5350.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., et al. (2013). Anthropogenic and natural radiative forcing. Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 658–740.

Philipona, R., Behrens, K., & Ruckstuhl, C. (2009). How declining aerosols and rising greenhouse gases forced rapid warming in Europe since the 1980s. *Geophysical Research Letters*, *36*, L02806. https://doi.org/10.1029/2008GL036350

Shindell, D. T., Lamarque, J.-F., Schulz, M., Flanner, M., Jiao, C., Chin, M., et al. (2013). Radiative forcing in the ACCMIP historical and future climate simulations. *Atmospheric Chemistry and Physics*, *13*, 2939–2974.

Smith, D. M., Booth, B. B. B., Dunstone, N. J., Eade, R., Hermanson, L., Jones, G. S., et al. (2016). Role of volcanic and anthropogenic aerosols in the recent global surface warming slowdown. *Nature Climate Change*, *10*, 936.

Smith, S. J., van Aardenne, J., Klimont, Z., Andres, R. J., Volke, A., & Delgado Abrias, S. (2011). Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmospheric Chemistry and Physics*, *3*, 1101–1116.

Storelvmo, T., Leirvik, T., Lohmann, U., Phillips, P. C. B., & Wild, M. (2016). Disentangling greenhouse warming and aerosol cooling to reveal Earth's climate sensitivity. *Nature Geoscience*, *4*, 286–289.

Taylor, K. E., Stouffer, R. J., & Meehl, G. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*, 485–498.

United Nations (2016). Paris Agreement (pp. 1–27). Paris: United Nations.

Wang, K., Ma, Q., Xiao, Y., & Wild, M. (2014). Urban impact of mean and trend of surface incident solar radiation. *Geophysical Research Letters*, *41*, 4664–4668. https://doi.org/10.1002/2014GL060201

Wilcox, L. J., Highwood, E. J., & Dunstone, N. J. (2013). The influence of anthropogenic aerosol on multi-decadal variations of historical global climate. *Environmental Research Letters*, *2*, 24033.

Wild, M. (2009). How well do IPCC-AR4/CMIP3 climate models simulate global dimming/brightening and twentieth-century daytime and nighttime warming? *Journal of Geophysical Research*, *114*, D00D11. https://doi.org/10.1029/2008JD011372

Wild, M. (2012). Enlightening global dimming and brightening. *Bulletin of the American Meteorological Society*, *1*, 27–37.

Wild, M. (2016). Decadal changes in radiative fluxes at land and ocean surfaces and their relevance for global warming. *Wiley Interdisciplinary Reviews: Climate Change*, *1*, 91–107.

Wild, M., Gilgen, H., Roesch, A., Ohmura, A., Long, C. N., Dutton, E. G., et al. (2005). From dimming to brightening: Decadal changes in solar radiation at Earth's surface. *Science*, *5723*, 847–850.
Wild, M., Ohmura, A., Schär, C., Müller, G., Folini, D., Schwarz, M., et al. (2017). The Global Energy Balance Archive (GEBA) version 2017: A database for worldwide measured surface energy fluxes. *Earth System Science Data*, *2*, 601–613.
Yang, S., Wang, X. L., & Wild, M. (2018). Homogenization and trend analysis of the 1958–2016 in situ solar radiation records in China. *Journal of Climate*, *31*, 4529–4541.