Understanding Northern Hemisphere land precipitation change

Submitted by

Joe Michael Osborne

to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Mathematics, May 2015.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

. . . . . .

Joe Osborne

Abstract

Water is key to life on Earth. The distribution and quantity of precipitation controls the availability of water, yet little is known about past changes in precipitation. This is especially true at regional scales. The land region of the Northern Hemisphere, especially the mid-latitude region, offers an ideal opportunity to compare and contrast output from climate models with our longest and most comprehensive precipitation observations. This thesis develops current understanding, with the aid of climate models, to attribute changes in global mean precipitation to known key forcings. Perhaps the most obvious feature of twentieth-century global mean precipitation change is a decrease in response to mid-twentieth-century aerosol forcing. Changes in historical precipitation for the land mean of two regions in the Northern Hemisphere are shown to strongly resemble changes in the global mean, due to a greater sensitivity to aerosol forcing than greenhouse gas forcing. This aerosol response is predictable across models, which offer an ideal resource to test this constraint due to their large range in aerosol forcing. However, this aerosol response is not evident in these key mid-latitude precipitation observations. Observed runoff changes, derived from river discharge measurements, also contrast with observed precipitation changes in this region. This contrast is a consequence of an obvious breakpoint in the runoff-precipitation relationship. An ensemble of land surface models that are driven with observed precipitation data fails to simulate this contrast and breakpoint. Combined, these two lines of evidence strongly suggest that Northern Hemisphere mid-latitude precipitation observations are unreliable, at least in the early twentieth century. It is expected that the true trend is disguised by inhomogeneities. This should be recognised in future research that is reliant on these data.

Acknowledgements

I am indebted to Hugo Lambert for his supervision throughout this PhD. I wish to thank him for giving me a chance in the first place and then showing incredible patience over the past few years. I fully appreciate the time he has spent helping me through my studies.

I would like to thank my viva examiners, Jim Haywood and Hayley Fowler, for taking the time to read my work and for a thorough but highly enjoyable discussion during the examination.

For the contribution of data, my thanks go to Mat Collins and Xuebin Zhang. I also thank my other co-authors for their data and feedback that contributed towards the runoff work in this thesis. Several other academics have offered help along the way, most notably Chris Ferro.

I would like to thank my friends from all five offices I have occupied, especially those who have provided advice, help, or even just a lunchtime escape from work. You have made my experience so much easier and enjoyable. I can only apologise for the smelly gym kit that so often resided under and around my desk.

I would also like to thank my parents, Kate, and Sarah for their love and support. My parents have seen opportunities for their children that did not exist for them. Also, Pumpkin and Obbi have provided frequent humour, for which I am enormously grateful.

Mostly, I would like to thank Lara for her unwavering love and support, and for never failing to put a smile on my face.

Contents

List of Figures

List of Tables

Abbreviations

Commonly used abbreviations are listed below. This list does not include model information (which are listed in separate tables throughout this work), but it does include frequently used datasets.

AOD: Aerosol optical depth

APS: Area per station

CMIP3: Coupled Model Intercomparison Project Phase 3

CMIP5: Coupled Model Intercomparison Project Phase 5

CRU: Climate Research Unit high-resolution (version 3.21) precipitation dataset

CRUNCEP: Merged CRU and National Centers for Environmental Prediction precipitation dataset

ECS: Equilibrium climate sensitivity

ERF: Effective radiative forcing

ERFaci: Effective radiative forcing due to aerosol-cloud interactions

ERFari: Effective radiative forcing due to aerosol-radiation interactions

ERFari+aci: Total effective radiative forcing due to aerosols

ESM: Earth system model

GCM: General circulation model

GHCN: Global Historical Climatology Network precipitation dataset

GHG: Greenhouse gas

GPCC: Global Precipitation Climatology Centre Full Data Reanalysis V6 precipitation dataset

GPCP: Global Precipitation Climatology Project precipitation dataset

HadCRUT4: Met Office Hadley Centre-Climate Research Unit version 4 temperature dataset

IPCC: Intergovernmental Panel on Climate Change

IPCC AR4: Intergovernmental Panel on Climate Change Fourth Assessment Report

IPCC AR5: Intergovernmental Panel on Climate Change Fifth Assessment Report

ITCZ: Intertropical convergence zone

LSM: Land surface model

LW: Longwave

NHML: Northern Hemisphere mid-latitude

NHT: Northern Hemisphere tropical

NMA: National meteorological agency

NWP: Numerical weather prediction

RFaci: Radiative forcing due to aerosol-cloud interactions

RFari: Radiative forcing due to aerosol-radiation interactions

SHEXT: Southern Hemisphere extratropical

SST: Sea surface temperature

SW: Shortwave

TOA: Top of atmosphere

TRENDY: A LSM intercomparison project

Zhang: Zhang precipitation dataset (based on GHCN; updated version of the Zhang et al. (2007) dataset)

Chapter 1

Introduction

Climate science has advanced enormously in recent years, aided by, for example, greater computational power and increased government funding. Not only has our scientific understanding of the climate system improved, but it is now possible to quantify the extent of change of many climate system components and also the human (anthropogenic) contribution towards these changes. For example, the recent (2013) Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) stated

> "...the science now shows with 95 percent certainty that human activity is the dominant cause of observed warming since the mid-20th century."

A common theme that emerges from such synthesis reports is that the most robust conclusions are made in relation to surface temperature changes. Even at the regional scale, the scientific community confidently details changes in temperature, which are largely controlled by the local energy balance (Shepherd, 2014). Despite these statements, such research is frequently revisited and revised owing to the ongoing heated debate between climate scientists and climate sceptics. This has been accentuated by the current temperature hiatus (e.g. Fyfe et al. (2013); Kosaka and Xie (2013); Cowtan and Way (2014)).

Precipitation is the key component of the hydrological cycle and changes in precipitation are expected to have more direct socio-economic impacts than changes in other climate variables (Gillett et al., 2004). From the most simplistic of viewpoints too much precipitation leads to flooding and too little precipitation leads to drought. But anthropogenic climate change means that some locations could see an increase in both drought and flooding as both the mean and distribution of precipitation changes i.e. less precipitation may fall over a year but in fewer, more intense, events. However, precipitation changes are more complex than temperature changes and precipitation studies are frequently limited by the quality of available observations.

At the global mean scale thermodynamic ideas place a constraint on precipitation changes (Chapter 2). But, of course, nobody lives in the global mean and at smaller scales – scales of interest for policy-makers – progress in scientific understanding of precipitation change has been slower. Regional precipitation is intrinsically related to atmospheric circulation and, as such, detailed dynamical understanding is needed to fully describe changes (Shepherd, 2014). Over land, physical processes are further complicated. Moisture availability introduces another constraint, so complex land-atmosphere interactions also influence precipitation (Koster et al., 2004). Orography has a strong influence on precipitation (Roe, 2005) and forcing from spatially heterogeneous short-lived forcing agents such as aerosols can shift precipitation both locally (Menon et al., 2002) and remotely (Chang et al., 2011). As a consequence, land precipitation is highly variable in space and time, even without human influences on the climate.

But a concentration on the Northern Hemisphere land area, in particular the mid-latitudes, offers a unique opportunity in our efforts to detail the mechanisms behind land precipitation change. This is because our most extensive – both temporally and spatially – in situ precipitation observations exist here. To fully understand past precipitation changes climate models need to be able to realistically replicate multiple aspects of observed change. Consistency with observations

is also a necessary constraint for an objective probabilistic forecast of precipitation changes (Allen and Ingram, 2002).

In this introduction methods of observing precipitation are considered, as well as how it is simulated in climate models. Limitations and opportunities of both are discussed. The current state of understanding of precipitation change is analysed. This includes changes in mean precipitation and changes in extreme precipitation, at a range of spatial scales. Finally, the wider hydrological cycle is discussed, with a focus on the land surface component, namely runoff, and how this can complement our understanding of precipitation change. The introduction finishes with an overview of the rest of the thesis.

1.1 Precipitation and atmospheric circulation

Precipitation is any form of water – liquid or solid – falling from the sky. It is a key part of the hydrological cycle. Much of the water vapour in the atmosphere has been evaporated from the oceans, but lakes and the land surface also provide a source of moisture (Trenberth et al., 2007). Precipitation forms following the condensation of water vapour and subsequent coalescence of cloud droplets to a point at which they become large enough to fall from the sky. The process of condensation occurs when an air parcel containing water vapour rises, cools and reaches saturation. Water vapour can be forced to rise in three main ways: 1) through baroclinic instability, associated with frontal precipitation, 2) through topography, associated with orographic precipitation and 3) through strong heating of the surface (or strong cooling aloft), associated with convective precipitation. The hydrological cycle is closely linked with the transfer of energy, both horizontally and vertically. The physical processes of evaporation and condensation take and release latent heat from and to the atmosphere respectively.

The distribution of precipitation across the Earth is controlled by atmospheric circulation, which arises from the rotation of the Earth, the equator receiving more

heat than the poles and the fact that solar heating largely occurs at the surface. The subsequent poleward redistribution of heat, combined with the Coriolis effect forms distinct patterns of precipitation. Perhaps the most robust feature is the intertropical convergence zone (ITCZ), which is found near the equator. The ITCZ forms in response to strong heating of the surface and crosses the equator biannually as the region of strongest heating shifts with the maximum in incoming solar radiation. The ITCZ is associated with very intense rainfall events and large annual mean rainfall values. The rising air of the ITCZ then moves polewards in the upper troposphere, before descending over the subtropics and returning to the equator via the surface, completing the Hadley cell tropical atmospheric circulation.

In the mid-latitudes there is an indirect meridional atmospheric circulation known as the Ferrel cell. It is indirect because it opposes the thermal gradient, with air rising in high latitudes (around 60–70°N), moving equatorwards and sinking in the subtropics (Vallis, 2006). This equatorward transport of heat is smaller than the poleward transport of heat arising from transient eddies. These eddies are found in baroclinic zones, where warm subtropical air, transported northwards at the surface, meets cold polar air. This redistribution of heat is associated with extratropical cyclones, a key source of precipitation over the mid-latitudes, particularly in the winter months when the temperature gradient is greatest. Such frontal precipitation is far less intense than precipitation from surface heating convective events, in general, although can persist for long periods of time and is locally enhanced by topography.

Because of the contrasting land masses in the Northern and Southern Hemispheres, the storm tracks differ. The abundance of continents in the Northern Hemisphere means that the prevailing westerly air-flow is more likely to be disrupted. However, two dominant storm tracks exist in the Northern Hemisphere, one in the Pacific Ocean and one in the Atlantic Ocean. With a prevailing west-to-southwesterly wind in the Northern Hemisphere mid-latitude (NHML) region,

west facing areas of the North American and European continents see some of the greatest annual mean precipitation totals over land in this region. Strong instability can be caused by cooling aloft, as well as heating at the surface. In winter, cold air advection over relatively warm waters can generate shallow convection and precipitation over land areas close to water-sources. Generally, where there is a source of moisture and instability convection may occur. Such convective events account for a greater percentage of summer precipitation in the mid-latitudes relative to winter.

A third circulation cell, the polar cell, extends from between 60° and 70° north and south to the poles. These are the smallest and weakest of the three main cells and, because cold air is able to hold little moisture, very little precipitation is seen here.

As well as this notable structure in precipitation patterns meridionally, there is also a lot of variability zonally. This is mostly due to the shape of the continents, which can give rise to areas of strong precipitation and with pronounced seasonality. These are known as monsoon systems. Because of the large heat content of the oceans relative to the land, the land warms much faster. With a land-sea temperature contrast often established in the spring season air rises over land, lowering the pressure and initiating low-level flow of moisture from nearby oceans. In the winter season, the land quickly cools and, combined with an ocean that is now warmer, the low-level flow reverses. This leads to winter dry and summer rainy seasons. Prominent monsoon systems include the South Asia Monsoon, the East Asian Monsoon, the Indo-Australia Monsoon and the West Africa Monsoon.

Precipitation variability can also arise due to intraseasonal variability in atmospheric circulation. The most obvious such mode of variability is the Madden-Julian oscillation (MJO) in the tropical atmosphere, which lasts between 30 and 60 days. The active phase consists of a band of enhanced convection and precipitation that propagates eastwards into the Pacific Ocean, following initiation in the Indian Ocean. A full dynamical understanding of MJO is lacking, but it has

been shown to modulate nearby monsoon rainfall and other modes of variability (Klingaman and Woolnough, 2014). It has even been shown to be a source of predictability for sub-seasonal to seasonal forecasting in the North Atlantic basin (Lin et al., 2010).

At longer timescales, annual or greater, precipitation variability can be driven by the El Niño Southern Oscillation (ENSO). Eastern equatorial Pacific sea surface temperature (SST) fluctuates – on a timescale of two to seven years – between anomalously cold (La Niña) and warm (El Niño) conditions. This can disrupt the Walker circulation, a zonally oriented component of the tropical Pacific atmospheric circulation, and shift areas favourable for atmospheric convection (Collins et al., 2010). Locally, South America sees large variability in precipitation but ENSO also influences precipitation in the mid-latitudes through global teleconnections (Toniazzo and Scaife, 2006). Around the North Atlantic basin, the North Atlantic Oscillation (NAO) is the dominant mode of variability and it has been shown to cause shifts towards wetter or drier conditions in parts of North America and Europe on multidecadal timescales (Hurrell, 1995)

1.2 Observing precipitation

1.2.1 In situ surface-based observations

Measurement of precipitation amounts is a long standing practice with the earliest record, the England and Wales Precipitation (EWP) series (Alexander and Jones, 2000), beginning in 1766. Recording changes in precipitation has long been of particular value to those working in agriculture and other industries. Land-based in situ precipitation measurements can be made simply, although not always accurately.

1.2.1.1 Basic principles

The most common approach to measurement of rainfall – initiated in the UK and followed widely elsewhere – has been for an observer to take a 0900 (local time) reading from a 5 inch storage rain gauge. The UK national meteorological agency (NMA), the Met Office, has tried to minimise the possibility of inhomogeneities arising in their long precipitation records. Because of the need for an observer to empty and measure the standard storage rain gauges daily, automatic instruments were introduced across the UK network in the 1980s and 1990s, typically employing the tipping bucket rain gauge (the bucket tips each time that 0.2 mm of rainfall is measured). While attempts have been made to keep the old storage gauges alongside tipping bucket gauges, in order to maintain a consistent record of measurements, this has not always been possible or practical. Indeed, where storage gauges have remained at a given site for multiple years, once daily readings may now only be taken weekly or even monthly.

1.2.1.2 Systematic errors

Measuring true precipitation amounts is difficult. There are a number of systematic errors, which, collectively, mean that measured precipitation is an underestimate of true precipitation. This phenomenon is known as undercatch. The most prominent source of undercatch is due to the wind effect, caused by wind eddies forming around the orifices of elevated rain gauges. Although it has been known about for a long time, the process is not fully understood. However, field studies, wind tunnel measurements and numerical simulations show the wind effect to be at a maximum in light rain and strong wind (e.g. Nešpor and Sevruk (1999)). Sevruk (1982) showed that, on average, the wind-induced error is 2–10% for rain.

Wind-induced undercatch becomes a particular problem when combining multiple precipitation datasets from different countries, as different countries employ different rain gauges as their standard. There are more than 50 types of national

standard gauges according to Sevruk and Klemm (1989). Further, the fact that the standard rain gauge has often been changed at least once, sometimes multiple times, within a country further compounds this problem. The Met Office shift towards automatic rain gauges is one such example. In the US (and other countries) a number of rain gauges are equipped with wind shields, although the number of shielded gauges at US National Weather Service (NWS) stations has reduced since the 1940s (Karl et al., 1993). A shielded gauge will record an amount closer to the true precipitation than an unshielded gauge at a given site. Yang et al. (1998) highlight that the combination of records from shielded and unshielded gauges, especially when the relative numbers vary temporally, can result in inhomogeneous precipitation time series.

At some stations the gauge may also be re-sited, either increasing or decreasing wind exposure and introducing an inhomogeneity. Sometimes more transient changes in surrounding buildings or vegetation are unavoidable. These can introduce trends in precipitation records not caused by true precipitation changes. Even an unchanged, unmoved rain gauge with a near-constant surrounding environment may see changes unrelated to true precipitation changes. This is because of changes in wind speed (Vautard et al., 2010) – forced either naturally or anthropogenically – and perhaps even changes in rain type/droplet size. For example, multiple studies have highlighted the suppression of drizzle due to aerosol pollution (e.g. Haywood and Boucher (2000); see Section 1.4).

Although not as significant as the wind effect, other systematic errors lead to undercatch of precipitation. These include wetting losses, which occur when rain collects on the inside walls of rain gauges but evaporates before being recorded. Losses can also arise due to evaporation of precipitation that has collected in a gauge. These are dependent on gauge-type and tend to be greatest for gauges without a funnel. Prior to the introduction of automated weather stations, both wetting and evaporation losses varied significantly with the frequency of measurements (Groisman and Rankova, 2001). Today, some manual gauges that were

previously emptied daily are only emptied weekly or even monthly, thus suffering from large evaporation losses.

Such systematic errors are even greater for solid precipitation than for liquid precipitation. The average wind-induced error increases from 2–10% for rain to 10–50% for snow (Sevruk, 1982). The Met Office report the water equivalent of freshly fallen snow as an alternative to gauge readings, where an observer is present. Where this is not the case, readings from rain gauges during snow events should be treated with caution. In high latitude countries, such as Canada and Russia, where snow is a common (or even the dominant) form of precipitation during the winter months, rain gauges are often adapted to improve the accuracy (reduce undercatch) of precipitation measurements in snow events. Generally, all the above sources of error in rainfall readings apply to snowfall readings but are enhanced in the latter. A consequence of this, in the absence of detailed corrections, is that a false wetting trend will appear in a warming world as the precipitation mix is dominated more by rain and less by snow (Rodda and Dixon, 2012).

1.2.1.3 Inhomogeneity studies

An ideal approach towards quality control would see corrections made for each individual station record. This would require a huge amount of metadata. Ideally this would include wind speed, temperature, precipitation type and precipitation intensity. Beyond meteorological variables, a detailed description of gauge type changes, siting changes, surrounding environment changes (exposure) and measurement frequency changes are also necessary to estimate true precipitation. Frequently such metadata are hard to find, undigitised or even non-existent. Individual stations are also subject to random errors, such as poor observer practice or even equipment malfunction at automated sites. These are near-impossible to account for but averaging over a large area tends to greatly reduce this random noise. In view of these difficulties, bias corrections are often made at the country

scale (e.g. Mekis and Vincent (2011)).

Precipitation observations in the former USSR have been particularly prone to changes affecting homogeneity. These include a change to a shielded gauge in the 1890s, a shift to two measurements per day in the 1930s, the introduction of new gauges in the 1940s, a wetting correction and move to four measurements per day in the 1960s and a subsequent move back to two measurements per day in the 1980s (Groisman et al., 1991; Groisman and Rankova, 2001). Groisman and Rankova (2001) focused on the wetting correction that was introduced in 1966-1967. This wetting correction was an effort to account for precipitation that reached the gauge collector or funnel but did not make the measuring glass. As a consequence all subsequent non-zero precipitation measurements were subject to a 0.2 mm correction for liquid precipitation and a 0.1 mm correction for solid precipitation. This correction lead to an increase in the number of precipitation days (a reading of zero was recorded prior to this correction when at least one drop was extracted from a gauge) and added 5–15% to annual totals across the former USSR. In the far north of Russia in winter months totals increase by up to 30%. However, it is difficult to adjust for this wetting correction because the number of trace events is difficult to determine. Because of this, some studies using Russian precipitation data consider just the period from 1966 to present (e.g. Wan et al. (2014)).

While these inhomogeneities are often easily identifiable as a jump discontinuity and have been corrected for, many are harder to identify. Recent work on Chinese observations found that bias corrections have changed precipitation trends in some regions (Ding et al., 2007). But while more complex time-space correction techniques are now being implemented (Stisen et al., 2012), adjustments are typically only made to the records of individual countries (Mekis and Vincent, 2011). Systematic and coordinated efforts have been made to identify and quantify biases in the construction of global surface temperature datasets (Morice et al., 2012). Such an effort for global land precipitation would require metadata on

as many individual stations as possible (Adam and Lettenmaier, 2003), together with appropriate correction techniques.

1.2.1.4 Gridded precipitation datasets

Data from multiple sources/countries have been combined to make century-long gridded land precipitation datasets, which have become a cornerstone of hydrometeorological research. Datasets vary in their construction but typically share the following common approaches. The datasets consist of regular grids covering the global land surface. Individual station data anomalies are calculated by subtracting the 1961-1990 climatology (average) on a monthly basis. The 1961-1990 period is commonly used because it is the period during which most observations are available. Since the 1990s the number of stations contributing towards gridded datasets has decreased dramatically, with NMAs trimming their observing networks due to lack of funding and sharing data less readily (New et al., 2001; Hegerl et al., 2014). Monthly station anomalies are then averaged within each grid box.

Gridded precipitation datasets are constructed independently but will all use some common station data. Some are constructed using a limited number of long-term homogenised records. The Global Historical Climatology Network (GHCN) dataset (Vose et al., 1992) combines data from 2,064 homogeneity adjusted precipitation stations (from the US, Canada and former USSR) with 20,590 unadjusted precipitation stations worldwide. Only stations that contribute at least 20 years of data in the 1961-1990 climatology period are included. These data were adapted in the Zhang et al. (2007) study to include only stations with at least 25 years of data during 1961-1990 and at least five years of data in every decade during 1950-1999. An updated version of this dataset (kindly provided by Xuebin Zhang) replaced Canadian data with the data based on the recent homogeneity study of Mekis and Vincent (2011). Both the GHCN and Zhang datasets are available on a $5^\circ \times 5^\circ$ grid, with spatial coverage limited to land grid boxes that contain at

least one station.

Other gridded precipitation datasets favour a greater number of short-term records. These include the Climate Research Unit (CRU) high-resolution precipitation dataset (Harris et al., 2014) (the CRU TS3.21 version is used here). The CRU dataset incorporates data from a greater number of sources. Although some NMAs that contribute data may have performed homogenisation on their records, no further bias corrections are applied. The Global Precipitation Climatology Centre (GPCC) Full Data Reanalysis V6 dataset uses considerably more stations than even the CRU dataset, combining long- and short-term stations to provide the best spatial coverage. This dataset places an emphasis on quantity of stations without addressing the quality of included records. Both the CRU and GPCC datasets are provided on a spatially interpolated $0.5° \times 0.5°$ grid offering complete global land coverage. However, both the CRU and GPCC datasets provide metadata to allow for analyses using the spatially incomplete, observations-only data.

Because different gridded precipitation datasets have different strengths and weaknesses, the above datasets will be considered together in analyses. It is expected that the most robust patterns of change at the continental scale will be evident across all datasets, allowing for robust conclusions. But these datasets will still contain inherent inhomogeneities, owing to their changing spatial coverage in time, in addition to the various systematic errors listed above. These inhomogeneities and biases may be shared, as different groups use many common records. The reliability of precipitation observations, particularly in the early twentieth century, will be considered when drawing conclusions.

1.2.2 Remote measurements

Traditional in situ measurements of precipitation offer reasonable temporal coverage – North America and Europe, in particular, are well observed back to

the beginning of the twentieth century – but are restricted to the land surface. Even when spatial coverage of gridded precipitation datasets reached a maximum around the 1970s, mountainous regions, where precipitation is highly variable (Adam et al., 2006), were still poorly represented.

Remotely sensed measurements from meteorological satellites are now made more readily than ever before, since their advent in the 1970s. They have the major advantage of offering near-complete global coverage (land and ocean). The traditional approach used passive retrievals in the visible or infrared frequencies to determine precipitation intensity based on the reflectivity of clouds or the temperature of cloud tops respectively. However, there are significant uncertainties in algorithms for converting radiometric measurements into precipitation (Ebert et al., 2007). Since the 1980s precipitation estimates from microwave sensors have been favoured, because they have been found to be more accurate (Ebert et al., 2007).

The most popular satellite-based gridded precipitation datasets, such as the the Global Precipitation Climatology Project (GPCP) dataset (Adler et al., 2003), merge uncertain (but high frequency of observation) infrared estimates from geosynchronous orbiting satellites with more reliable (but low frequency of observation) microwave estimates from polar orbiting satellites. These satellite-based products are then adjusted by surface-based gauge observations (Adler et al., 2003; Huffman et al., 2007). Such merged algorithms are still developed with the tropics in mind, because the relationship between radiometric measurements and precipitation is most robust here (mainly because geosynchronous orbiting satellites struggle to "see" the polar regions. Therefore, the accuracy of satellite-based precipitation datasets decreases polewards of 50° (Hegerl et al., 2014). Accuracy also decreases over land areas, relative to oceans, and notably over snow or ice.

So while satellite-derived precipitation estimates appeal because of their vastly improved spatial coverage over in situ surface based measurements, they are arguably *currently* more susceptible to inaccuracies and inhomogeneities than the

traditional measurement approach. They also suffer from brevity. At the decadal scale it can be hard for a climate signal to emerge from natural variability, especially at smaller spatial scales. Here, to understand precipitation changes over the Northern Hemisphere land area it is necessary to use century scale records. This allows second half of the twentieth century changes, when anthropogenic climate forcing became more complicated but was notable for a strengthening and subsequent weakening in anthropogenic aerosol forcing (Lamarque et al., 2010; Wild, 2012), to be placed in context with changes from the early twentieth century, when greenhouse gases (GHGs) were the dominant anthropogenic climate forcing.

1.2.3 Reanalyses

Atmospheric reanalyses also offer complete global coverage. Estimates of numerous meteorological variables are made by assimilating historical observations into the latest numerical weather prediction (NWP) models. Reanalyses are affected by the changing amount of assimilated observational data in time. Precipitation derived from reanalyses such as ERA-Interim (Dee et al., 2011), which often performs the best in evaluation studies (e.g. Lorenz and Kunstmann (2012); Lindsay et al. (2014)), is particularly limited and is not recommended for climate trend analyses and long-term water budget studies (Lorenz and Kunstmann, 2012). Many of the more established reanalysis datasets are also restricted to the second half of the twentieth century.

The problems with reliability and short records have been recently tackled with the development of a new dataset, the Twentieth Century Reanalysis (20CR). This reanalysis assimilates surface synoptic observations only and extends temporal coverage back to 1871 (Compo et al., 2011). But this means that precipitation is simply a product of the model and not an assimilated variable (true of all reanalyses). Kalnay et al. (1996) classified reanalysis output variables depending on how influenced they are by observations and the NWP model. Precipitation was assigned to the least reliable of these classes, reflecting that it is only indirectly

related to the assimilated observations.

1.3 Changes in precipitation

Precipitation change has been studied at a range of spatial and temporal scales. Generally, the signal-to-noise ratio improves at greater scales. But even at large scales great care must be taken to disentangle signals arising from natural variability and long-term forcing. Natural shifts in precipitation climatology can persist for decades due to modes of variability such as ENSO and NAO. Changes in annual mean precipitation may disguise contrasting changes in seasonal precipitation, which are of greater importance socio-economically.

As well as changes in mean precipitation, changes in the probability distribution of precipitation intensity are important, particularly with regard to changes in drought and flood frequency. For example, an increase in mean precipitation for a region may mask an increase in extreme precipitation but a decrease in light or moderate precipitation, a scenario that could increase occurrences of both droughts and floods. The type of precipitation is also important, particularly for flood management and water resource planning. Here, the most robust features of precipitation change are considered.

1.3.1 Global mean

Atmospheric moisture is tightly coupled to global temperatures. The Clausius-Clapeyron relation implies that specific humidity increases roughly exponentially with temperature at about 7% K^{-1} (assuming constant relative humidity). This is a robust feature that has been found in column integrated water vapour measurements from the Special Sensor Microwave/Imager (SSM/I) (Wentz and Schabel, 2000) (a satellite-mounted instrument), ERA-Interim reanalysis (Dee et al., 2011) and surface observations (Willett et al., 2008). From this one could naively as-

sume that the increase in global mean precipitation with temperature is similar. But Mitchell et al. (1987) showed that the increase in evaporation and precipitation simulated by an early general circulation model (GCM) was only 2.8% K^{-1}. Increases in global mean precipitation of 2–3% per K of warming have since been found in several generations of GCMs (e.g. Allen and Ingram (2002); Held and Soden (2006); Lambert and Webb (2008); Thorpe and Andrews (2014)). This is because changes in global mean precipitation are constrained by the tropospheric energy budget, rather than the availability of moisture, as discussed in Section 2.1.

Due to the absence of long-term (century scale) observations over the oceans it is hard to verify this energetic constraint with real-world values. Using a record length of 20 years (1987-2006) Wentz et al. (2007) suggested that SSM/I measured global mean precipitation increased by 6.7 $\pm$ 3.5% K^{-1} (following a slight downward adjustment by Lambert et al. (2008) after using more suitable statistics). But the short length of the record means that natural variations in climate can dominate, giving large precipitation changes per K warming. Another satellite-based estimate (from the GPCP dataset), for the 1988-2008 period, is much lower at 2.8 $\pm$ 0.9% K^{-1} (Allan et al., 2014). On smaller time scales aerosols can also influence the magnitude of the global mean precipitation increase per degree of warming. This is because scattering aerosols, absorbing aerosols and GHGs have different direct effects[1] on precipitation – an influence independent of the forced change in surface temperature – on the tropospheric energy budget (Chapter 2).

[1]The term "direct effect" can mean different things within different climate science fields. Hereinafter, unless otherwise stated, it refers to the direct effect on precipitation and *not* the aerosol direct effect. As is traditional in the literature, the direct and indirect effects of aerosols discussed in this thesis (see Section 1.4) refer specifically to radiative effects. In contrast, precipitation direct effects concern tropospheric energetics, specifically changes in tropospheric latent heating associated with changes in precipitation directly due to any given climate forcing (see Section 2.1.2). Other changes, mediated through changes in surface air temperature, are known as temperature dependent effects (see Section 2.1.1).

1.3.2 Spatial patterns

The spatial pattern of precipitation change in response to GHG-driven global warming is expected to be complex, but certain features have been proposed and revisited in recent years. The most obvious type of response in precipitation minus evaporation (P - E) proposed is that wet regions get wetter and dry regions get drier (Held and Soden, 2006; Chou et al., 2009). Increases in water vapour content mean that there will be additional moisture convergence in the convergence zones, fuelled by increased moisture divergence in the descent zones (the subtropics), where evaporation rates increase (assuming that the spatial patterns of circulation remain the same). But increases in moisture content are greater than increases in precipitation. There must be a balance between the global mean rate of precipitation and the moisture transport into the free troposphere, which is a product of boundary layer moisture content and mass exchange between the boundary layer and the free atmosphere (Vecchi et al., 2006). Therefore, the tropical circulation must weaken.

Changing spatial patterns of precipitation are best described as a combination of thermodynamic and dynamic changes. Chadwick et al. (2013) showed that the net spatial pattern of precipitation change in the tropics is dominated by spatial changes in convection, which results from, for example, changes in SST gradients. The importance of such dynamic changes suggests that the wet gets wetter, dry gets drier response is not hugely useful for describing precipitation change in the tropics, even over the oceans. A robust feature of both Coupled Model Intercomparison Project Phase 3 (CMIP3) and Coupled Model Intercomparison Project Phase 5 (CMIP5) (these intercomparison projects formed the basis of climate model analyses in IPCC AR4 and IPCC AR5 respectively) GCMs in response to future warming is the poleward expansion of the Hadley cell. This leads to a poleward expansion of the subtropical dry zone. As a consequence, the most robust precipitation declines are found poleward of the subtropical precipitation minima (Scheff and Frierson, 2012).

A similar, but opposite, pattern has been proposed in the mid-latitudes. Under a fixed atmospheric circulation assumption the mid-latitudes are expected to see more precipitation in a warmer world (Held and Soden, 2006; Seager et al., 2010). But the precipitation maxima associated with the mid-latitude storm tracks may also move polewards, although CMIP5 models simulate this far less robustly than the older CMIP3 models. The P - E response over land is particularly complicated because it is often moisture-limited rather than energy-limited. Greve et al. (2014) found that only about 11% of the global land area has followed the wet get wetter, dry get drier pattern since 1948.

In addition to changes in the spatial pattern driven by homogeneous GHG forcing are changes driven by heterogeneous aerosol forcing, a key focus of this thesis. Aerosols can cause regional precipitation changes both locally and remotely (Shindell et al., 2012), as discussed in more detail in Section 1.4.4 and in Chapter 3. While changes in regional precipitation are complex, the availability of century-long surface-based precipitation records over parts of the Northern Hemisphere land mass (see Section 1.2) at least offers an opportunity to detail robust real-world changes here and, with the aid of GCMs, attribute the reason (aerosols or otherwise) behind such changes. Likewise, aerosols are assumed to have had a detectable influence on global mean precipitation (Chapter 2) and on runoff (Chapter 4).

1.3.3 Extreme precipitation

The most extreme precipitation events occur when nearly all available atmospheric moisture precipitates from a system. Therefore, increases in precipitation extremes are likely to be constrained by the increase in water vapour content of 7% K^{-1} (Trenberth et al., 2003). This is generally verified by observed daily precipitation extremes, with Westra et al. (2013) reporting increases in annual maximum daily extreme of 5.9–7.7% per K change in global mean temperature. However, the rate of change in daily precipitation extremes would appear to be meridionally

variable, although the physical mechanisms behind these variations are not always apparent.

O'Gorman (2012) used GCM predictions of extremes constrained by observations of interannual variability in the current climate – GCMs by themselves have difficulty in reliably simulating precipitation extremes (see Section 1.5) – to show that extreme (99.9th percentile) daily precipitation in the tropics increases at a rate of 10% per K of surface warming. While most of the contribution towards increases in extreme tropical precipitation is expected to come from increases in water vapour content, a dynamical mechanism in the form of increased horizontal moisture flux convergence could also play a role in explaining the enhanced sensitivity to global temperature (O'Gorman and Schneider, 2009). Alternatively, the sparse nature of precipitation observations in the tropical region may not lend itself towards accurate quantification of changes in precipitation extremes (Westra et al., 2013). The mid-latitude sensitivity of extreme daily precipitation to global temperature has been found to be lower at about 5% K^{-1} (O'Gorman and Schneider, 2009; Zhang et al., 2013), closer to that implied by the Clausius-Clapeyron relation. Recent detection and attribution studies have found a human fingerprint on the intensification of extreme precipitation over Northern Hemisphere land (Min et al., 2011; Zhang et al., 2013).

Sub-daily precipitation extremes are even more difficult to study, largely because of the limitations of the observed precipitation record – older, non-automated rain gauges were typically only emptied and recorded at 12- to 24-hour intervals (see Section 1.2.1.1) – and the fact that sub-daily precipitation extremes cannot be realistically simulated by GCMs. However, Berg et al. (2013) illustrated the rich resource that is available in radar measurements, which can provide five-minute rainfall estimates. A common theme from extreme sub-daily precipitation studies – typically limited to short observational time periods – is that these short duration extreme events are intensifying more rapidly than daily precipitation extreme events. This is known as super-Clausius-Clapeyron scaling and occurs at

local atmospheric temperatures in excess of about 12–15°C (Lenderink and van Meijgaard, 2008). Berg et al. (2013) advanced recent work by using three-hourly synoptic observations over Germany to separate precipitation into convective and stratiform types depending on the observed clouds. This addressed suggestions that super-Clausius-Clapeyron scaling may just be a consequence of changes in the precipitation type. The authors found that the intensity of extreme sub-daily stratiform precipitation increased with temperature at Clausius-Clapeyron rates, but this rate was exceeded for extreme sub-daily convective precipitation.

Although intense sub-daily duration precipitation events have sometimes caused localised flooding and significant economic losses in the UK (e.g. Warren et al. (2014)), multi-day precipitation events can cause the greatest socio-economic disruption by leading to severe and prolonged flooding. Fowler and Kilsby (2003) (updated in Jones et al. (2013)) showed that there have been increases in spring and autumn extreme precipitation events at a range of multi-day durations between 1961 and 2009, with increases at 5- and 10-day durations across the north of the UK most prominent. Meanwhile, 5- and 10-day duration winter maxima have increased in intensity, particularly over Scotland, with precipitation increases of up to 50mm for such events. Hannaford and Marsh (2008) found corresponding increases in high streamflow frequency and magnitude, most notably during the winter season and in the north and west of the UK. Such increases in multi-day duration heavy precipitation events may have contributed towards the exceptional floods across much of the UK in the winter of 2013/14 (Huntingford et al., 2014), although drivers of precipitation increases are still uncertain.

1.4 Aerosols

Aerosols are microscopic liquid or solid particles suspended in the atmosphere. They can originate from natural sources, including volcanic emissions, sea salt and mineral dust, as well as anthropogenic sources. Anthropogenic air pollution can

arise from activities such as biomass burning, so has not simply been restricted to the twentieth century. This type of pollution was, however, confined to a very small spatial scale and the associated climate effects were negligible. But in the NHMLs – particularly North America and Europe – in the years following the Second World War a period of rapid industrialisation saw a large spike in anthropogenic air pollution (Stern, 2006). Anthropogenic emissions of sulphur and black carbon increased from the 1950s to 1980s, before a subsequent decrease in recent decades (Stern, 2006; Smith et al., 2011). This pattern is consistent with the implementation of clean air policies introduced to prevent serious health effects.

In the past 20 years Asian aerosol emissions have increased. This has been proposed as one possible reason behind the global surface temperature hiatus since the late 1990s. But Kühn et al. (2014) showed that the effect of these sulphur and black carbon emission changes on the regional and global energy balance between 1996 and 2010 was small. This is because cooling caused by sulphate aerosols has been opposed by warming by black carbon aerosols. Meanwhile, a reduction in sulphate aerosols across Europe has changed cloudiness and caused a positive radiative global mean forcing overall (Kühn et al., 2014).

1.4.1 Aerosol-radiation interactions

The aerosol direct effect, referred to as the radiative effect due to aerosol-radiation interactions (REari) in IPCC AR5, is the change in radiative flux caused by the combined scattering and absorption of radiation by anthropogenic and natural aerosols (Boucher et al., 2013). The main anthropogenic scattering aerosols are sulphate, nitrate and organic carbon, while the main anthropogenic absorbing aerosol is black carbon (Figure 1.1). It is still difficult to reliably quantify REari, with known difficulties in both observation and simulation. The climate effects of aerosols depend on their distribution, as well as their optical properties. The distribution is further complicated by the short tropospheric lifetime of aerosols

(typically of the order of a week), the heterogeneity of their sources and the dependence on the synoptic meteorological conditions. The fraction of scattered and absorbed radiation by an aerosol depends on the size distribution, refractive index and mixing state.

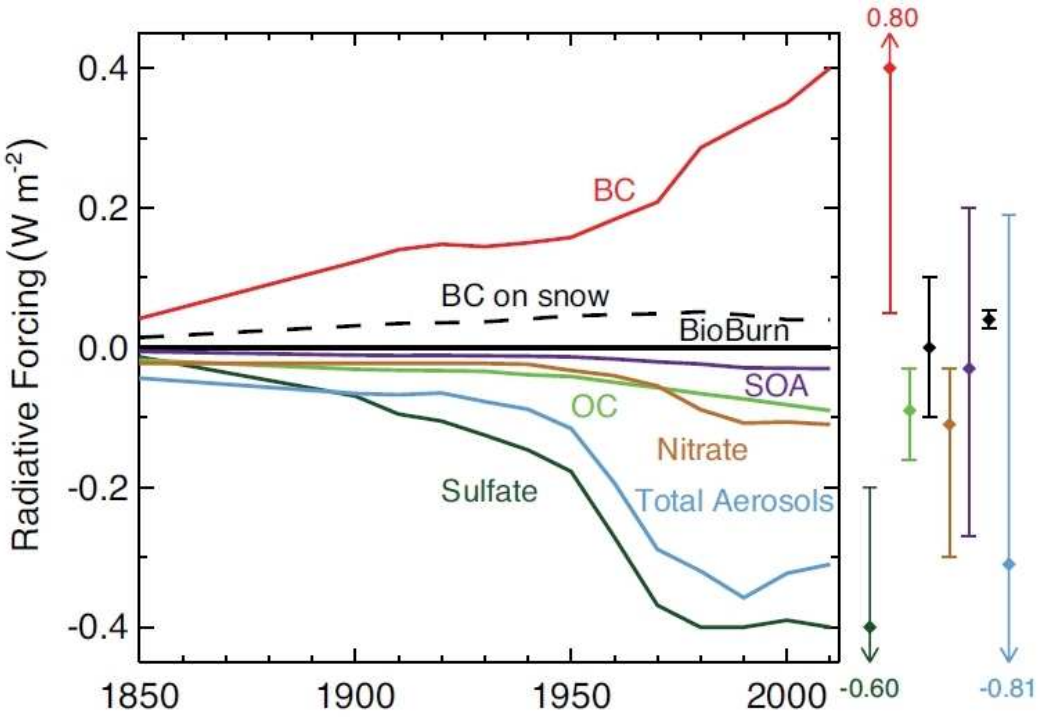

Figure 1.1: Time series of global annual mean top of atmosphere (TOA) radiative forcing due to aerosol-radiation interactions and black carbon (BC) on snow and ice. OC is organic carbon and SOA is secondary organic aerosol. The vertical error bars represent 5–95% uncertainty ranges and the values next to the error bars indicate where the uncertainties extend beyond the scale. From Figure 8.8 of Myhre et al. (2013).

Measurement of these key aerosol properties is achieved through laboratory and field studies. Results from such studies are aiding the improved representation of aerosol properties in models. Global aerosol models are initialised using analyses from atmospheric data assimilations and anthropogenic emissions from standard inventories (e.g. Dentener et al. (2006)). Aerosol abundance is simulated in time, with radiative transfer models used to calculate the associated radiative forcing. High spectral resolution, line-by-line models are considered the benchmark but are too computationally expensive to be used in standard global aerosol models. Heavily parameterised radiative transfer schemes are used instead, introducing another source of uncertainty in model-based calculations of the aerosol direct effect.

Observational data are also used to determine the REari. Measurements of aerosol optical depth (AOD), a measure of the column-integrated aerosol amount, are made from ground-based networks and sensors attached to satellites. One of the most widely used instruments in studies is the Moderate Resolution Imaging Spectrometer (MODIS) (e.g. Bellouin et al. (2008); Myhre (2009)). The difference between the total AOD and the AOD due to natural aerosols is the AOD due to anthropogenic aerosols (Bellouin et al., 2008). Other retrievals from the MODIS sensor, together with an algorithm, allow for the determination of the fraction of total AOD due to various aerosol types. Estimates vary significantly between satellite products, owing to difficulties with calibration, whether the REari is assumed to be zero or not in cloudy conditions and treatment of surface reflectivity. Satellite retrievals at visible wavelengths using passive sensors (e.g. MODIS) are not possible over highly reflective surfaces, such as deserts and snow- and ice-covered regions, and additionally cannot make retrievals at high latitudes during winter owing to the lack of solar insolation. However, UV wavelength channels can be used to make AOD retrievals over some bright surfaces (Hsu et al., 2006).

The radiative forcing due to aerosol-radiation interactions (RFari) is still uncertain, but is more robust in AR5 than in previous IPCC assessment reports. Part of this improvement is due to greater agreement between estimates of RFari from global aerosol models and observations. Myhre (2009) demonstrated improved consistency between estimates of RFari from these two sources, when adjustments are made to the observations given forcing in cloudy regions and pre-industrial aerosol concentrations. Given these improvements and other recent intercomparison projects the most likely global mean RFari is -0.35 [-0.85 to +0.15] Wm^{-2} (values in brackets represent the 5–95% uncertainty range) (Boucher et al., 2013). This is less negative than in IPCC AR4 because the positive forcing from absorbing species is now believed to be greater than before (Bond et al., 2013). The effective radiative forcing due to aerosol-radiation interactions (ERFari), which accounts for rapid adjustments (see Section 2.2), is slightly more negative at -0.45 [-0.95 to +0.05] Wm^{-2} (Boucher et al., 2013).

1.4.2 Aerosol-cloud interactions

Anthropogenic aerosols also contribute a radiative forcing due to aerosol-cloud interactions (RFaci). These aerosol-cloud interactions, also known as aerosol indirect effects, are summarised schematically in Figure 1.2. Because of the very short time between aerosol-cloud interactions and subsequent cloud adjustments it is difficult to estimate a RFaci alone. Instead, the effective radiative forcing due to aerosol-cloud interactions (ERFaci) is considered. The ERFaci includes microphysical modifications to the cloud albedo (the first aerosol indirect effect) (Twomey, 1977), as well as changes in the cloud lifetime (the second aerosol indirect effect) (Albrecht, 1989).

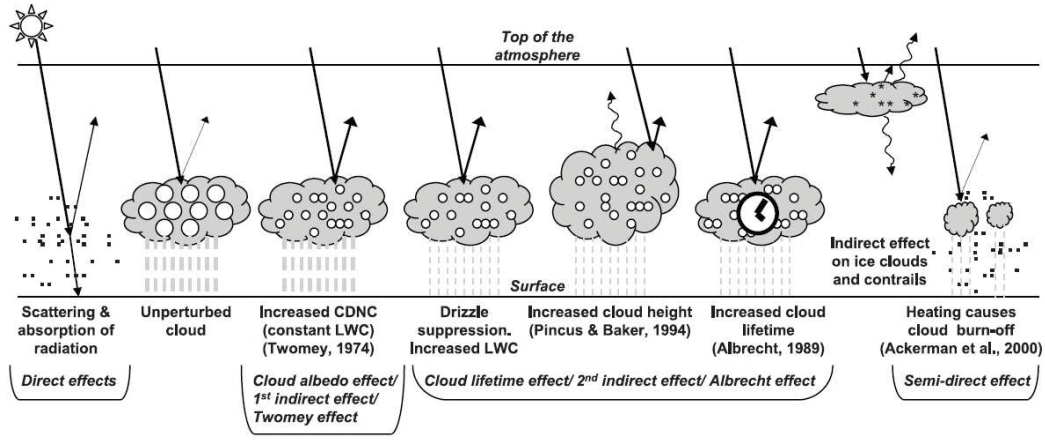

Figure 1.2: Schematic of the aerosol indirect effects. CDNC is cloud droplet number concentration and LWC refers to the liquid water content. From Figure 2.10 of Forster et al. (2007), which was adapted from Figure 1 of Haywood and Boucher (2000).

These microphysical modifications arise from the fact that aerosols serve as cloud condensation nuclei (CCN) and ice nuclei (IN), on which cloud droplets can activate. Droplet activation occurs when an air parcel is lifted and cooled adiabatically to a level at which saturation occurs, known as the lifting condensation level. Therefore, activation is dependent on the environmental lapse rate (Malavelle et al., 2014). But the type, size and number of aerosols also affect activation. A greater aerosol number concentration means a greater cloud droplet number concentration. Therefore, for a constant liquid water content, the droplet size will decrease. The result is an increase in cloud reflectance. Even more complex and poorly quantified is the effect on cloud lifetime. The well-established idea, first

proposed by Albrecht (1989), is that the reduction in cloud droplet size means that clouds can contain more liquid water before precipitation, thus increasing the cloud lifetime (and even the cloud thickness).

Earlier IPCC assessment reports considered the cloud albedo and cloud lifetime effects to be complementary, but improved process level understanding has revealed the full complexity of cloud responses to changing aerosol concentrations. It is now clear that the radiative forcing due to aerosol-cloud interactions is sensitive to local environmental conditions and that these two aerosol indirect effects can oppose one another. But these compensating effects are typically not accounted for in GCMs. The lack of robust observational evidence to drive and validate model development is also a hindrance (Rosenfeld et al., 2014). Although observational evidence from aircraft measurements and satellite-derived measurements verify the physical basis behind the albedo effect, these two lines of evidence disagree on the magnitude of this effect. Generally, satellites are limited in that they are unable to "see" the bottom of a cloud, the region most important for aerosol nucleation.

Many of the CMIP5 GCMs incorporate a representation of the first aerosol indirect effect and several even include the second aerosol indirect effect (see Table 1 of Wilcox et al. (2013)). But the microphysical nature of aerosol-cloud interactions means that they are occurring at sub-grid scales. Consequently, their representation in GCMs is reliant on numerous parameterisations. Recent developments include parameterisations of aerosol activation (Ghan et al., 2011), but these still depend heavily on other unresolved processes, such as updraft velocity. In fact, even NWP models running at a horizontal resolution of 1 km fail to resolve individual convective cells (Malavelle et al., 2014). New super-parameterisation schemes are being used to bridge the gap between cloud-resolving models, which can resolve individual clouds but are computationally expensive, and GCMs.

In IPCC AR5 the total effective radiative forcing due to aerosols (ERFari+aci) is given greater attention than the ERFaci alone. The ERFaci can be calculated as

the residual between ERFari+aci and ERFari if one assumes linearity of ERFari and ERFaci. They are unlikely to be additive though.

1.4.3 Total aerosol effect

The combined ERFari+aci (global mean change in the net radiation at the top of atmosphere (TOA) from pre-industrial to present day) is given as -0.9 [-1.9 to -0.1] Wm^{-2} based on expert judgement (Boucher et al., 2013). This is a compromise between the less negative ERFari+aci of -0.85 [-0.93 to -0.45] Wm^{-2} from satellite observations and the more negative ERFari+aci of -1.38 [-1.68 to -0.81] Wm^{-2} from GCM studies. The latest GCMs that include a more complete representation of aerosol-cloud interactions tend to have a less negative ERFari+aci than older GCMs. This pattern is reflected in the shift to a smaller magnitude aerosol forcing in IPCC AR5 relative to IPCC AR4, the best estimate previously being -1.2 Wm^{-2} (Forster et al., 2007). Despite this downward revision, aerosols remain the most significant forcing after GHGs (Figure 1.3).

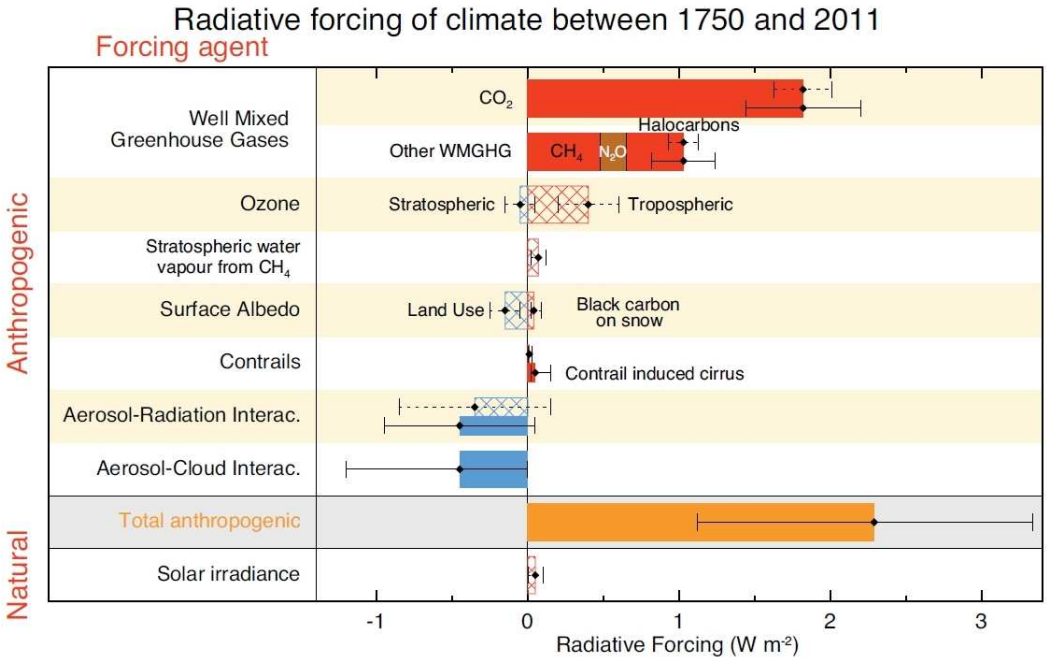

Figure 1.3: Global mean TOA forcing for 2011 relative to pre-industrial (1750). Hatched bars represent radiative forcing and solid bars represent effective radiative forcing. 5–95% uncertainty ranges are shown (dotted lines for radiative forcing and solid lines for effective radiative forcing). From Figure 8.15 of Myhre et al. (2013).

1.4.4 Aerosols and precipitation change

Aerosol-cloud interactions are likely to have little influence on the net atmospheric heating rate and thus are not expected to have a direct effect on regional or global scale precipitation. But the combined ERFari+aci is expected to influence precipitation patterns at a range of spatial scales. The heterogeneous nature of aerosol radiative forcing means that surface cooling (and strong atmospheric warming in the case of black carbon aerosol) means that both vertical and horizontal temperature gradients can be altered. Cooling of the surface and/or heating of the atmosphere reduces the environmental lapse rate and increases atmospheric stability. This has been proposed as a key mechanism behind the drying of the South Asian summer monsoon (Ramanathan et al., 2005). However, Bollasina et al. (2011) suggested that changes in atmospheric stability play a secondary role to the more important changes in circulation, whereby South Asian aerosol forcing instigates an energy imbalance between the Northern and Southern Hemispheres – seen in observed changes in the meridional SST gradient – and there is a compensating slowdown of the tropical meridional overturning circulation.

Zhao et al. (2006) showed a strong relationship between recent precipitation decreases over eastern central China and increases in atmospheric stability, derived from meteorological soundings. Increasing anthropogenic aerosols were proposed as the reason behind these changes in stability. Evidence has also been found for a local effect on precipitation by US anthropogenic aerosols. Leibensperger et al. (2012) found a general decrease in evaporation, due to reduced incoming solar radiation, with a clear decrease in downwind precipitation in response. Although mid-twentieth-century aerosol increases were characterised by precipitation decreases, the south central US was found to be an exception. This was attributed to forced circulation changes that increased the flow of moisture into the south central US from the Gulf of Mexico.

Significant surface temperature gradient changes and associated atmospheric circulation changes are a consequence of the heterogeneous nature of aerosol forcing.

This means that precipitation over a region can be more strongly influenced by remote aerosol forcing than local aerosol forcing (Shindell et al., 2012). The aerosol emissions from the NHML region around the 1960s induced an anomalous temperature gradient (Friedman et al., 2013). This is believed to have been the cause behind a southward shift in the tropical rain belt seen in models and observations (Chang et al., 2011; Hwang et al., 2013) (Chapter 3). Asymmetric forcing induced by stratospheric aerosols (rather than tropospheric aerosols) in a geoengineering or volcanic scenario was also found to shift the tropical rain belt, with it moving away from the hemisphere that sees greater cooling (Haywood et al., 2013). Aerosol forcing also influences global mean precipitation (Chapter 2) and the land surface component of the hydrological cycle (Chapter 4).

1.5 Modelling precipitation

GCMs are computer models that simulate the Earth system using numerical methods. The core equations governing the fields of thermodynamics, fluid dynamics and radiation (although heavily parameterised) allow for a simulation of observable processes across multiple components of the climate system. The earliest 3D climate models in the 1970s (e.g. Manabe and Wetherald (1975)) constituted an atmospheric GCM coupled to a simple "slab" ocean model that omitted ocean dynamics. Since then computing speeds have increased by roughly a factor of a million. The benefits of this increased speed have been divided between producing models of greater complexity (more components and processes), running longer simulations and improving spatial resolution.

Components in current GCMs include atmosphere, land surface, ocean and sea ice, interactive aerosols, atmospheric chemistry, dynamic vegetation and interactive carbon cycle. These components have been developed offline before being coupled into the climate model, whereby they can interact with the other components. Despite increasing spatial resolution, there are still some processes that

occur at sub-grid scale. These processes need to be parameterised. Parameterisations are guided by fundamental physical principles and make use of the large scale simulated conditions in a given grid box. Precipitation is one such variable that depends on small scale processes, with even microphysical interactions of importance. Therefore, simulation of precipitation relies heavily on parameterisations.

1.5.1 Precipitation parameterisation

Both CMIP3 and CMIP5 GCMs have separate moist convective and large scale precipitation parameterisations to simulate convective and stratiform precipitation types. Parameterisations vary considerably in their complexity. Moist convection is split into two categories: 1) deep convection, which can reach the tropopause, and 2) shallow convection, which is typically capped at 500 hPa. The main features of convection that need to be parameterised include triggering/activation mechanisms, closure assumptions and the cloud model (detailing the vertical distribution of heating, moistening and momentum changes) (Dai, 2006).

The earliest and simplest convective schemes calculated precipitation as the fraction of moisture convergence that was required to heat the atmosphere and relax the temperature profile towards the moist adiabatic lapse rate, in the presence of conditional instability (Kuo, 1965). But convection does not simply consume water at the rate of supply (instead it consumes energy – see Section 1.3.1) and modern schemes are far more complex. Many, such as the Gregory and Rowntree (1990) scheme used in the Met Office Hadley Centre GCMs, are based on a bulk mass flux approach. This uses one single "bulk" cloud model, derived by summing over an ensemble of convective clouds with differing characteristics (Plant, 2010). Parameterisations deal with a number of detailed processes, such as entrainment and detrainment (high entrainment rates inhibit deep convection by mixing the parcel with the environment), changes of phase and evaporation of

falling precipitation.

The core equations in GCMs, combined with the resolution of GCMs, allow for the natural formation of fronts (boundaries separating air masses of different temperature) through baroclinic instability, which bring large scale precipitation. However, precipitation associated with these fronts still requires a parameterisation scheme. The more basic representations in CMIP3 models simply produced stratiform precipitation whenever the local relative humidity was supersaturated. More detailed schemes (e.g. Smith (1990) – which still forms the basis of the stratiform cloud scheme used in the Met Office Hadley Centre GCMs) also account for cloud amount and water content, as well as changes in atmospheric stability brought about by phase changes. More recent schemes have separate representations of maritime and continental stratiform clouds (Lohmann and Roeckner, 1996). Improvements in atmospheric aerosol modelling have led to more realistic representations of the number of activated aerosols (dependent on aerosol properties and vertical velocity) (Donner et al., 2011), adding further detail to precipitation parameterisation schemes.

1.5.2 Quality of simulation

Models typically perform well in reproducing the broad observed spatial pattern, as well as the annual cycle of precipitation, while year-to-year internal variability is generally realistic (Dai, 2006). Some exceptions have persisted through model developments, such as the double-ITCZ problem. Mehran et al. (2014) showed that CMIP5 models still tend to replicate mean precipitation over arid regions and other subcontinental regions poorly, while most models produce too much precipitation over complex topography. At smaller time scales, generations of models have struggled to accurately represent the diurnal cycle of precipitation, with deep convection developing too early in the day (often before midday). Another long standing problem is that GCMs tend to underestimate the intensity and frequency of heavy precipitation, but overestimate the intensity and frequency

of light precipitation. These problems and others are more extensively detailed by Stephens et al. (2010). Despite these persistent biases in model simulated precipitation, it is expected that large scale continental mean precipitation (scales that are studied here) is reasonably simulated across the CMIP5 ensemble, as errors over subcontinental regions tend to cancel.

1.5.3 CMIP5 vs. CMIP3

Models have continued to evolve between CMIP3 and CMIP5. Knutti and Sedlacek (2013) estimated that there has been an increase by about a factor of 60 in computational capacity between CMIP3 and CMIP5, with a factor of 2–4 of this going into improved model resolution and the increase in simulated years accounting for a factor of 2–3. Therefore, a factor of roughly 3–10 has gone into improving the representation of complex processes within climate models. This improvement in the representation of complex processes has seen a move towards Earth system models (ESMs) (Taylor et al., 2012). ESMs differ from GCMs in that they include interactive carbon cycles where, through simulation of complex biological and chemical processes, fluxes of carbon between the various climate system components are accounted for. Although many of the CMIP5 models fall under the ESM definition, hereinafter the CMIP5 multi-model ensemble is simply referred to as the CMIP5 GCMs.

Many recent developments have centred around cloud microphysics, particularly aerosol-cloud interactions (see Section 1.4.2). Several models now include a representation of aerosol indirect effects and have generally improved or incorporated parameterisations of microphysical processes. The atmospheric component of GFDL-CM3 now also incorporates interactive atmospheric chemistry, so that aerosols are prescribed at the emission stage and allowed to evolve according to the current climate (Donner et al., 2011). The HadGEM2-ES model includes an improved Gregory and Rowntree (1990) mass flux convective scheme after a revision of the treatment of convective detrainment, with NWP tests showing significant

improvements in model performance (Derbyshire et al., 2011).

An important consideration in much of this work and other studies that consider multi-models ensembles is the degree to which models can be considered truly independent. It is known that there are interdependencies among CMIP5 GCMs. For example, some institutions develop multiple models that often only differ in their ocean components. Also, some models from separate institutions use the same parameterisations, code and, in a few cases, entire components (the Australian ACCESS models are based on the HadGEM2 atmosphere) (Knutti et al., 2013). Regardless, in most studies models are given equal weighting and as a result the spread is not a good estimate of uncertainty, which is typically underestimated. Also, the spread in a finite number of models is clearly not representative of the spread of all possible models. This complicates analyses, particularly uncertainty quantification. But CMIP5 generally performs much better than CMIP3 in comparisons with observations (e.g. Knutti and Sedlacek (2013), but there are numerous such studies).

1.6 The wider hydrological cycle

The global water and energy cycles are closely related through radiative fluxes, as well as latent heat and sensible heat fluxes. Precipitation is just one component of the wider hydrological cycle. Components are likely to respond in different ways in a warming world, but will also share some common trends and variability. The oceans are the main water reservoir and source of evaporation globally. Approximately 90% of evaporated ocean water precipitates back into the oceans (Trenberth et al., 2007). The remaining ∼10% is transported in the form of water vapour to the land and is returned as continental freshwater runoff after being precipitated, thus maintaining a freshwater balance in the oceans. The return to the oceans is mostly concentrated at the mouths of rivers. Land precipitation is also driven by evaporation and transpiration – collectively known as evapo-

transpiration – from the land surface itself, which (globally) still accounts for the majority of land precipitation.

Because changes in the individual components– water vapour, cloud, evapotranspiration, soil moisture, permafrost, runoff/streamflow – can strongly impact societies, great efforts have been made to observe their changes, understand their changes and predict their future changes. The easiest component to measure historically has been precipitation (see Section 1.2), but skill in accurately observing other components has advanced greatly in recent years with the advent of remote sensing and focussed field campaigns. Over the last 10–15 years a large number of worldwide streamflow measurements have been collated to create historical monthly streamflow datasets covering much of the land surface (Dai et al., 2009). Assuming that evapotranspiration is equal to the residual of precipitation and runoff, historical changes in the three main components of the land water cycle can be quantified.

The partitioning of precipitation into runoff and evapotranspiration is complicated, largely because it is a function of aridity, the balance between water demand (potential evapotranspiration) and supply (precipitation), which varies across latitude belts (Budyko, 1974). But streamflow trends and internannual-to-decadal variations are expected to be driven by precipitation (Milliman et al., 2008; Dai et al., 2009). Therefore streamflow, a measure of integrated runoff within a river catchment, can be used to validate precipitation changes (Chapter 4).

1.7 Research aims and objectives

This thesis brings together several lines of evidence – from key observations, coupled model intercomparision projects and land surface models (LSMs) – to develop an improved understanding of historical changes in large scale precipitation and the wider hydrological cycle. It seeks to bypass some of the flaws GCMs have in simulating aspects of the hydrological cycle and instead identify robust features

that are understood. At regional scales the agreement between observations and model simulations is tested and the implications for any disagreement considered.

The key aims of this thesis are:

1. To determine the impact of contrasting direct effects of different climate forcings, particularly GHGs and aerosols, on twentieth-century global mean precipitation change. This will be achieved by:

   (a) revisiting the tropospheric energy budget constraint on global mean precipitation and considering its implication for transient, historical changes,

   (b) developing an idealised framework to elucidate the effect of aerosol forcing on historical precipitation changes,

   (c) testing and validating this framework using the latest GCMs.

2. To explore the extent to which a similar aerosol signature might be expected to be seen and is seen in regional land mean precipitation changes in both models and observations. This aim will be met by:

   (a) determining whether regional land mean changes scale with global mean changes,

   (b) comparing and contrasting model simulated and observed precipitation responses to aerosol forcing in these regions.

3. To determine whether observed regional precipitation changes are consistent with observed changes in regional runoff by:

   (a) using a statistical model to test for unusual behaviour in the observed runoff-precipitation relationship,

   (b) using data from LSMs driven with observed precipitation to simulate runoff.

1.8 Thesis structure and key findings

The rest of this thesis details new results, split into three separate chapters. It finishes with a conclusion of overall findings,

Chapter 2 looks at global mean precipitation changes and begins by reviewing the tropospheric energy budget constraint that forms the basis of most research at this scale. Developing previous ideas it is shown that global mean precipitation changes can be successfully split into temperature-dependent changes and temperature-independent changes. In keeping with previous research it is shown that these two components can be separated using an idealised, abrupt CO_2 concentration increase experiment and a simple linear regression technique. The sensitivity of precipitation change to temperature change is found to be similar to estimates made using older GCMs. There is little spread across CMIP5 models. This has been found previously using CMIP3 models and reduced complexity models. Although temperature independent, or direct effect, precipitation changes can only be estimated for changing CO_2 concentrations using this techniques, these are found to be significant and oppose temperature-dependent changes.

Next, a well-established approach for diagnosing longwave (LW) and shortwave (SW) radiative forcing time series that relies on calculating the climate feedback parameters from the idealised, abrupt CO_2 increase experiments is applied. This chapter then continues by developing existing literature to show that model simulated transient twentieth-century global mean precipitation changes can be detailed using 1) an understanding of the strength of past climate forcings, and 2) an understanding of their respective direct effects. This takes a simplistic approach by considering just the GHG and sulphate aerosol components of twentieth-century forcing. It is argued that if GHG and sulphate aerosol forcing are modelled as a linear increase and a step change respectively, then the global mean precipitation-temperature relationship can be thought of as two straight lines "offset" from one another. This offset is caused by the step change in sulphate aerosol.

This pattern is then found to be a real feature in the CMIP5 model simulated global mean precipitation-temperature relationship. It is hypothesised that the magnitude of the offset is dependent on a model's strength of aerosol forcing. Through defining methods for determining the magnitude of these variables it can be seen that this is a robust feature of twentieth-century global mean precipitation change in CMIP5 models, with the largest offsets seen in models with the strongest (most negative) aerosol forcing. This is also robust across intercomparison projects, even though the CMIP3 models do not sample such a large range in aerosol forcing as the CMIP5 models.

The fact that the range of aerosol/SW forcing is so much greater in CMIP5 than in CMIP3 raises an interesting aside that has not been satisfactorily tackled by the existing literature. Why do CMIP5 models still simulate twentieth-century global mean temperature changes so well? A narrow range of end of twentieth-century temperature changes across models can result from a correlation between net forcing and climate feedback/sensitivity. Interestingly this correlation is found to be significant in CMIP5 models, just as it was for CMIP3 models, despite predictions to the contrary due to the larger range in SW forcing. Here, the prevalence of this correlation is attributed to a remarkable anti-correlation between SW forcing and LW forcing, which acts to converge net forcing in CMIP5 models.

This chapter finishes by exploring the aerosol offset in more detail and giving more evidence for why it is such a predictable feature. This requires a return to the very idealised framework that only considers GHG and sulphate aerosol forcing. Using this approach the size of the offset is shown to be dependent on the strength of sulphate aerosol forcing and a coefficient that determines the ratio of atmosphere LW to TOA LW forcing. The latter is shown to be fairly constant across models and model intercomparison projects.

Building on these ideas that apply at the global mean scale, **Chapter 3** explores whether regional land mean precipitation changes, for which twentieth-century observations are available, show obvious features that can be associated with

aerosol forcing. An offset is shown to be evident in a CMIP5 model when considering both the NHML land mean precipitation and Northern Hemisphere tropical (NHT) land mean precipitation relationships with global mean temperature, in realistic twentieth-century experiments. Such a feature is absent from idealised simulations that consider just GHG or aerosol forcing alone.

It would appear that CMIP5 models are able to capture wetting and drying trends in observed NHT land mean precipitation (which have been previously attributed to remote NHML aerosol forcing), but there exists a clear mismatch in observed and modelled NHML land mean precipitation. In CMIP5 models changes in both NHT land mean precipitation and NHML land mean precipitation scale broadly with changes in the global mean. And it is expected that a reduction in precipitation should be evident in response to local aerosol forcing. The size of the precipitation offset in models (in both land regions) is dependent on the strength of NHML aerosol forcing. Further it is shown that the observed NHT land mean precipitation response is in keeping with estimates of real-world twentieth-century NHML aerosol forcing, but the NHML land mean precipitation offset is of opposite sign to what is expected.

So why do our longest, most comprehensive precipitation observations not show such a robust aerosol signal? It is suggested that this missing aerosol response in these observations may be a consequence of inhomogeneities in the observed record. But this is purely speculative at this stage and the possibility that CMIP5 models incorrectly simulate twentieth-century NHML land mean precipitation change cannot be ruled out.

Chapter 4 looks to further explore this idea that inhomogeneities may exist in mid-latitude precipitation observations. This is achieved by using complementary runoff observations, a key component of the hydrological cycle. It is argued that – in the absence of net changes in water storage due to human activities – for a land region consisting of one or more river catchments precipitation is equal to the sum of runoff and evapotranspiration. Although humans can have multiple (direct and

indirect) human influences on the hydrological cycle, that affect the partitioning of precipitation into evapotranspiration and runoff, twentieth-century changes in runoff are expected to be linear with twentieth-century changes in precipitation.

Using a set of six LSMs that are driven with the same observed gridded precipitation dataset, the interannual, decadal and multidecadal changes in NHML mean runoff are shown to closely match those in NHML land mean precipitation, although whole-century runoff trends are less positive than whole-century precipitation trends, indicating that these LSMs simulate increases in evapotranspiration over this period. In contrast to the significant positive trend in NHML land mean precipitation, there is a non-significant (small) positive trend in NHML mean runoff. Closer inspection of the runoff and precipitation records shows that the difference in trends can be attributed to a breakpoint in the runoff-precipitation relationship around the 1930s.

Using a statistical model to test for changes in this relationship, a 1930s breakpoint is found to be a robust feature. The work of the previous chapter and the evidence from the LSMs – where no such breakpoint is detected – strongly suggest that this is a consequence of an inhomogeneity in the NHML land mean precipitation time series, rather than the NHML runoff time series. Applying the breakpoint method at continental scales, the North American continent is seen to be the most likely source of this breakpoint. Further lines of evidence are shown to suggest that this breakpoint is not due to changing spatial coverage in time (sampling different climatic regions) or dam and reservoir construction, or associated increases in irrigation.

Applying an adjustment to the NHML land mean precipitation time series, given the magnitude of the breakpoint, somewhat accounts for (more than halves) the discrepancy between the precipitation response expected and that observed. This lends support to ideas proposed in the previous chapter. But the fact that this technique does not fully account for this discrepancy suggests that further work is required to bridge this gap in understanding.

Chapter 5 presents the main conclusions and discusses some of the limitations to the work. It also presents ideas for future research and summarises the contributions to the research field.

1.9 The contribution of papers to the research field

A shorter version of Chapter 2 has been published as a paper in *Nature Climate Change* (Osborne and Lambert, 2014). This study was designed and developed with my supervisor Hugo Lambert, but I performed the analysis and wrote the manuscript. The key conclusion was that twentieth-century NHML precipitation observations are missing an expected aerosol response. This finding has implications for research areas that rely on these key precipitation observations, such as model validation and offline hydrological modelling that uses precipitation observations as a crucial input. Although previous work had suggested that early-twentieth-century precipitation observations may be unreliable, many studies still use data from this period. This work concludes by suggesting that inhomogeneities may be prevalent in precipitation observations and future work should consider the reason behind these. However, the possibility that model simulations of precipitation are unreliable is also not dismissed.

The work presented in Osborne et al. (2015) (published in *Journal of Hydrometeorology*) serves as a follow up to the work of Osborne and Lambert (2014), attempting to either support or rebuke the conclusions of the earlier paper. This work grew from discussions with my supervisor about whether other closely linked hydrological observations were in keeping or not with precipitation observations. The most obvious observations to use were river discharge records, which can be related to runoff. This would appear to be the first such study to discuss changes in mid-latitude runoff and precipitation in tandem and, further, use the former to consolidate ideas concerning the latter.

I was the lead author of this large collaborative study, with the methodology devised and results discussed with both Hugo Lambert and Margriet Groenendijk. A total of eight co-authors performed LSM experiments and lent their output to this paper, as well as giving feedback that improved the study. The findings presented showed that the observed NHML land mean runoff-precipitation relationship contained a breakpoint that was not detected in the LSMs, with observed NHML mean runoff more in keeping with what was expected given mid-twentieth-century aerosol forcing. This provided further evidence that the precipitation observations contain inhomogeneities. This will hopefully encourage future research into the source of inhomogeneities and the spatial coverage of observing networks necessary to reliably detail changes in precipitation.

Chapter 2

Global mean precipitation

Determining changes and drivers of changes in global mean precipitation is an important research topic. Although the value of such research may not be immediately apparent to, say, policy-makers it can form the basis of research into changes at smaller spatial scales. Here, the tropospheric energy budget constraint, introduced in Section 1.3.1, is considered in more detail. This constraint, coupled with the evolution of key climate forcings through the twentieth century, can help explain the pattern of global mean precipitation changes in GCMs. A key aspect of twentieth-century climate change is aerosol forcing. By introducing methods for measuring the strength of aerosol forcing and the magnitude of the resulting precipitation response, a robust framework is built. The chapter continues with a consideration of the key features of models, which determine their temperature and, in turn, precipitation characteristics. Finally, it is asked whether and why the precipitation response to aerosol forcing should be so robust in view of intermodel differences.

2.1 The tropospheric energy budget

Changes in the overall intensity of the global mean hydrological cycle are constrained by the tropospheric energy budget (Mitchell et al., 1987; Allen and In-

Model	Institute	Country	Ensemble size
ACCESS1-0	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM)	Australia	1
bcc-csm1-1	Beijing Climate Center, China Meteorological Administration	China	3
CanESM2	Canadian Centre for Climate Modelling and Analysis	Canada	5
CCSM4	National Center for Atmospheric Research	USA	6
CNRM-CM5	Centre National de Recherches Météorologique	France	10
CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	Australia	10
FGOALS-s2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences	China	1
GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory	USA	5
GFDL-ESM2G			1
GFDL-ESM2M			1
GISS-E2-H	NASA Goddard Institute for Space Studies	USA	5
GISS-E2-R			5
HadGEM2-ES	Met Office Hadley Centre	UK	4
inmcm4	Institute for Numerical Mathematics	Russia	1
IPSL-CM5A-LR	Institut Pierre-Simon Laplace	France	5
IPSL-CM5A-MR			3
IPSL-CM5B-LR			1
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	Japan	4
MIROC-ESM			3
MPI-ESM-LR	Max-Planck-Institut für Meteorologie	Germany	3
MPI-ESM-P			2
MRI-CGCM3	Meteorological Research Institute	Japan	3
NorESM1-M	Norwegian Climate Centre	Norway	3

Table 2.1: The 23 CMIP5 models used throughout this thesis (smaller model ensembles are sometimes used depending on availability of experiments, number of runs and data access), associated modelling groups and countries. Also shown is the ensemble size of the historical (all forcings) experiment. The mean across ensemble members is taken for models (where multiple initial condition runs were made) throughout this work.

Model	Institute	Country
BCCR-BCM2.0	Bjerknes Centre for Climate Research	Norway
CCSM3	National Center for Atmospheric Research	USA
CGCM3.1	Canadian Centre for Climate Modelling and Analysis	Canada
CNRM-CM3	Centre National de Recherches Météorologique	France
CSIRO-Mk3.0	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	Australia
ECHAM4	Istituto Nazionale di Geofisica e Vulcanologia	Italy
ECHAM5	Max-Planck-Institut für Meteorologie	Germany
ECHO-G	Meteorological Institute of the University of Bonn, Institute of KMA, and Model and Data Group	Germany/Korea
GFDL-CM2.0 GFDL-CM2.1	NOAA Geophysical Fluid Dynamics Laboratory	USA
GISS-EH GISS-ER	NASA Goddard Institute for Space Studies	USA
INM-CM3.0	Institute for Numerical Mathematics	Russia
IPSL-CM4	Institut Pierre-Simon Laplace	France
MIROC3.2(hires) MIROC3.2(medres)	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change	Japan
MRI-CGCM2.3.2	Meteorological Research Institute	Japan

Table 2.2: The 17 CMIP3 models used throughout this chapter, associated modelling groups and countries. Unlike with CMIP5 models, only one ensemble member (run 1) from each model is considered. Although multiple members were available for precipitation and temperature variables for some models, forcing time series provided (see Section 2.2.3) were calculated from just one ensemble member.

gram, 2002). The troposphere has a small heat capacity. It can be assumed to be in equilibrium on time scales of a year or more. Therefore, changes in tropospheric latent heating associated with changes in global mean precipitation ($L\Delta P$), where $L = 2.5 \times 10^6$ J kg^{-1} (or $\frac{1}{L} \approx \frac{1}{29}$ W^{-1}m^2 mm day^{-1} in equivalent precipitation units) is the latent heat of vapourisation, are balanced by changes in tropospheric cooling (ΔR), through changes in the net radiative and sensible heat fluxes, such that

$$L\Delta P = \Delta R. \tag{2.1}$$

Assuming that global mean tropospheric temperature change is consistent with global mean surface air temperature (ΔT), and thus maintains a moist adiabatic lapse rate, an increase in the latter will lead to a tropospheric cooling that allows an increase in global mean precipitation. Therefore, global mean precipitation change is a function of global mean surface air temperature. Typically, global mean precipitation responds to changes in global mean surface air temperature at a rate of 1–3% K^{-1} (Held and Soden, 2006; Lambert and Webb, 2008) in GCMs, including CMIP5 models (Thorpe and Andrews, 2014). It has been noted though (e.g. Allan et al. (2014)) that the transient ΔP response to ΔT is muted. This is because a climate forcing that produces a change in ΔT can also directly change tropospheric cooling. Linearising ΔR equation (2.1) becomes

$$L\Delta P \simeq k_T \Delta T + \Delta R_A, \tag{2.2}$$

where $k_T \Delta T$ represents the global mean surface air temperature-dependent tropospheric cooling and ΔR_A represents cooling due to forcings and is independent of ΔT.

Therefore, different radiative forcings that cause a given change in global temperature can produce different changes in precipitation, owing to their different direct effects (Lambert and Allen, 2009). As a result, twentieth-century precipitation change is a function of surface temperature change and the direct effect of forcing agents driving the surface temperature change. The consequences of this on the

twentieth century precipitation-temperature relationship are explored further in Section 2.3.

Disentangling the competing components of global mean precipitation change, $k_T\Delta T$ and ΔR_A, is difficult in a transient run, particularly one in which multiple forcing agents are included. However, using idealised experiments, in particular the CMIP5 quadrupled CO_2 experiments (abrupt4×CO_2), it is possible to calculate the magnitude of these components (Figure 2.1). Such experiments also provide an illustrative tool to explain the separate components.

At the start of these experiments the CO_2 concentration is instantaneously quadrupled and then held constant. This leads to an initial decrease in precipitation as the atmosphere rapidly adjusts to the change in TOA radiative forcing and ΔR_A in equation (2.2) is strongly negative i.e. there is a net decrease in tropospheric cooling due to increased CO_2 concentrations. This is because the troposphere becomes more opaque to LW radiation, which is then emitted to space from higher, colder levels. Following this initial decrease there is a quasi-linear increase in global mean precipitation with the more slowly responding - due to the large heat capacity of the oceans - global mean surface air temperature, the $k_T\Delta T$ component. For these reasons $k_T\Delta T$ and ΔR_A are often referred to as the "slow" and "fast" responses respectively (Allen and Ingram, 2002; Andrews et al., 2010; O'Gorman et al., 2012).

2.1.1 Temperature-dependent precipitation changes

Using ordinary least squares regression of $L\Delta P$ against ΔT, k_T ($\mathrm{Wm^{-2}\ K^{-1}}$) – the sensitivity of tropospheric cooling to ΔT – is the gradient and ΔR_A is the y-intercept in step experiments (Figure 2.1). The ΔT dependent effects of climate forcings on precipitation have been shown to produce increases of 1.3–3.2 $\mathrm{Wm^{-2}\ K^{-1}}$ (or 1.4–3.4% $\mathrm{K^{-1}}$) using perturbed physics and CMIP3-era models (Lambert and Webb, 2008). Here, the multi-model mean from 23 CMIP5 models (Table

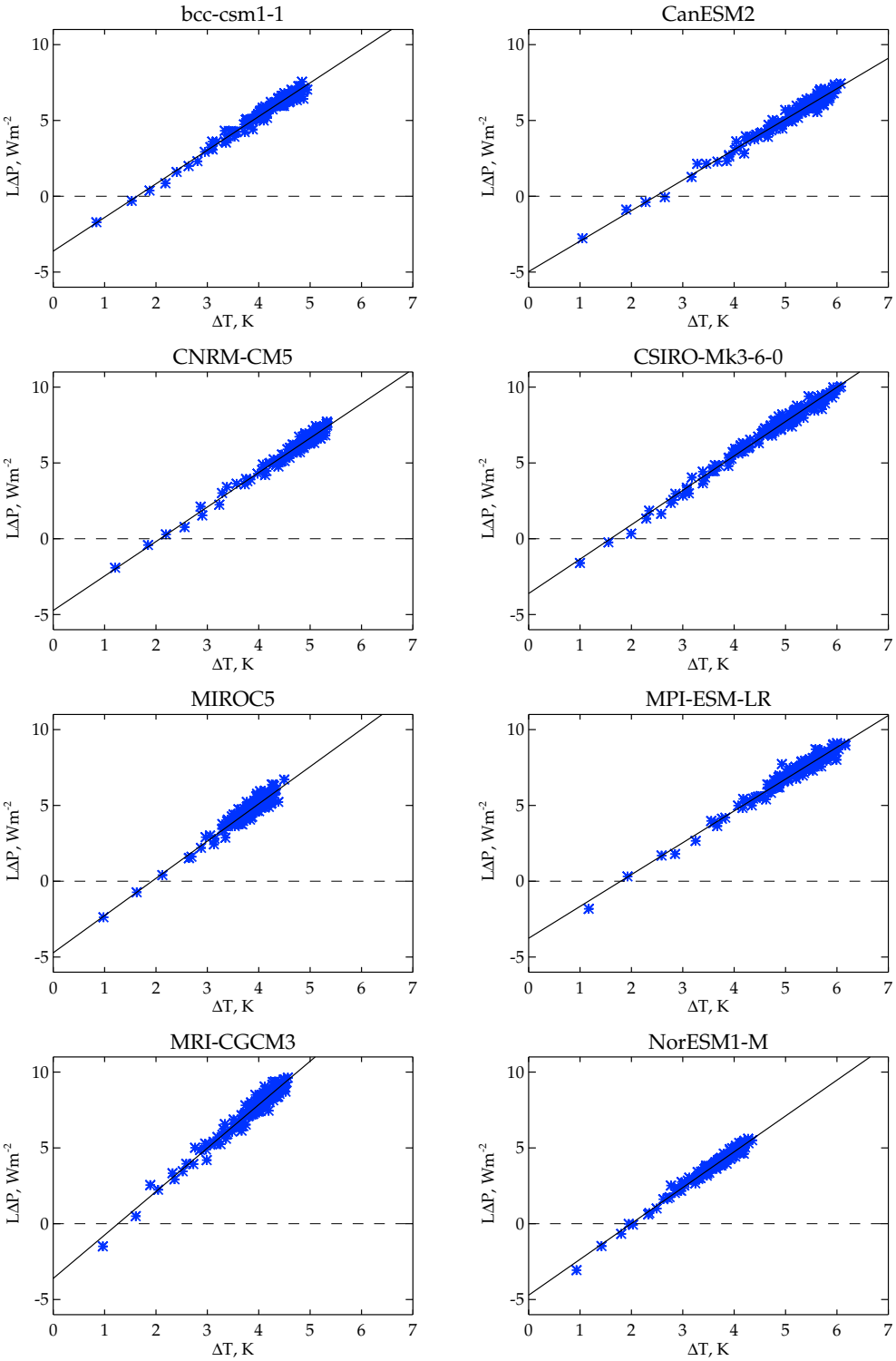

Figure 2.1: Relationships between changes in tropospheric latent heating associated with changes in global mean precipitation ($L\Delta P$) and global mean surface air temperature (ΔT), after an instantaneous quadrupling of CO_2 in the idealised abrupt4×CO_2 experiments. Data points are global annual means. Lines represent ordinary least squares regression fits to 150 years of data. The intercept at $\Delta T = 0$ is equal to ΔR_A (the fast response can be scaled to 2×CO_2 conditions by halving) and the gradient is equal to k_T. Anomalies are with respect to the mean global annual mean of each model's preindustrial control simulation. A subset of 8 of the 23 CMIP5 models is shown.

	k_T ($\mathrm{Wm^{-2}\ K^{-1}}$)	ΔR_A ($\mathrm{Wm^{-2}}$)
ACCESS1-0	1.85 ± 0.03	-2.68 ± 0.08
bcc-csm1-1	2.22 ± 0.05	-1.85 ± 0.10
CanESM2	2.01 ± 0.05	-2.48 ± 0.12
CCSM4	2.43 ± 0.05	-2.60 ± 0.11
CNRM-CM5	2.27 ± 0.05	-2.34 ± 0.12
CSIRO-Mk3-6-0	2.27 ± 0.04	-1.75 ± 0.09
FGOALS-s2	1.94 ± 0.06	-1.17 ± 0.15
GFDL-CM3	2.25 ± 0.05	-2.30 ± 0.12
GFDL-ESM2G	1.89 ± 0.09	-1.88 ± 0.15
GFDL-ESM2M	2.04 ± 0.09	-2.19 ± 0.15
GISS-E2-H	2.52 ± 0.05	-3.05 ± 0.09
GISS-E2-R	2.32 ± 0.17	-2.56 ± 0.26
HadGEM2-ES	1.89 ± 0.03	-2.68 ± 0.09
inmcm4	2.43 ± 0.09	-1.84 ± 0.13
IPSL-CM5A-LR	2.54 ± 0.04	-2.67 ± 0.11
IPSL-CM5A-MR	2.55 ± 0.04	-2.65 ± 0.11
IPSL-CM5B-LR	2.28 ± 0.06	-2.37 ± 0.12
MIROC5	2.46 ± 0.10	-2.51 ± 0.19
MIROC-ESM	2.12 ± 0.04	-1.96 ± 0.13
MPI-ESM-LR	2.10 ± 0.05	-1.91 ± 0.13
MPI-ESM-P	2.20 ± 0.06	-2.11 ± 0.14
MRI-CGCM3	2.87 ± 0.08	-1.85 ± 0.15
NorESM1-M	2.36 ± 0.05	-2.39 ± 0.09
Multi-model mean	**2.25 ± 0.41**	**-2.25 ± 0.70**

Table 2.3: Estimates of k_T and ΔR_A from 23 CMIP5 models (Table 2.1), using the abrupt4×CO_2 experiments. Estimates of ΔR_A have been scaled to 2×CO_2 conditions by halving. For individual models errors are the error in the ordinary least squares regression fit and for the multi-model mean the error is calculated from the range of best estimates across the 23 models (all 5–95% uncertainty ranges).

2.1) is 2.25 ± 0.41 $\mathrm{Wm^{-2}\ K^{-1}}$ (Table 2.3). The range of k_T values appears to be quite small and a robust feature of GCMs, independent of their complexity – from models with a simple slab ocean (thermodynamics-only) to fully-coupled atmosphere-ocean models.

Lambert and Webb (2008) asked why this range should be so small when compared to the range in equilibrium climate sensitivity (ECS) (see Section 2.2.1)? It is known that various climate feedbacks affect the value of k_T (Previdi, 2010). The dominant contributor to the magnitude of the temperature feedback is the Planck feedback component (associated with vertically uniform warming), although it shows little scatter across GCMs. Meanwhile the lapse rate feedback (associated with vertically non-uniform warming) and the water vapour feedback (mostly due to reduced LW cooling) tend to oppose each other. SW cloud feedbacks are very uncertain and mostly responsible for the large scatter in ECS estimates from

GCMs. Importantly though, SW cloud feedbacks have little effect on the tropospheric energy budget, because they mainly concern the scattering of radiation in the atmosphere, as opposed to absorption.

Andrews et al. (2010) showed that the ΔT dependent precipitation response is independent of the climate change mechanism. It is also frequently assumed to be independent of the type of experiment, such as an idealised instantaneous forcing experiment, or a more realistic transient twentieth-century experiment. But Good et al. (2012) suggested that k_T is also a function of the amount of CO_2, with smaller values of k_T found at a higher CO_2 concentration. Overall, the global mean precipitation response to this component would appear to be well understood and lends confidence to statements that global mean precipitation trends are muted with respect to trends of atmospheric moisture (see Section 1.3.1). It also seems very unlikely that LW cloud feedbacks – one of the remaining sources of scatter in k_T estimates – can be proposed as a mechanism to reconcile these differing trends (Lambert et al., 2014).

2.1.2 Direct effect precipitation changes

Past and future changes in GCM simulated global mean precipitation are dominated by the $k_T \Delta T$ effect (Lambert and Webb, 2008), although there is far more uncertainty in the magnitude and even the sign of the direct effects of forcing agents. Idealised experiments have shown that the direct effect of changing CO_2 concentrations on $L\Delta P$ is significant (Figure 2.1 and Table 2.3), but small or even consistent with zero for solar forcing (Lambert and Faull, 2007; Andrews et al., 2009). More generally, any forcing that absorbs LW and/or SW radiation in the troposphere and creates an imbalance in TOA minus surface forcing will contribute a direct effect (Frieler et al., 2011). As with solar forcing, sulphate aerosols will have little direct effect, because these concern the scattering of SW radiation (Andrews et al., 2010; Kvalevåg et al., 2013).

Conversely, black carbon aerosols have a significant direct effect because they strongly absorb solar radiation in the troposphere. In fact, the direct effect of black carbon forcing leads to a strong reduction in precipitation, which can actually oppose the temperature dependent effect to such an extent that total precipitation decreases (Andrews et al., 2010; Kvalevåg et al., 2013). However, the precipitation response to the direct effect of black carbon aerosol forcing is somewhat complicated and is dependent on the level in the troposphere that the black carbon aerosol is found (Ming et al., 2010; Kvalevåg et al., 2013). Greater precipitation reductions due to the direct effect are found when the aerosol burden is placed at a greater altitude in idealised experiments.

2.2 Diagnosing radiative forcing

Radiative forcing is defined in IPCC AR5 as the "change in the net, downward minus upward, radiative flux (expressed in Wm^{-2}) at the tropopause or TOA, due to a change in an external driver of climate change." These external drivers, also called radiatively active agents, include changes in GHG concentrations and changes in aerosol concentrations. It is defined net downward (LW + SW) so that an increase in radiative forcing leads to a warming and a decrease in radiative forcing leads to a cooling. Radiative forcings are frequently used in climate science as they offer a means of comparing the influence on surface temperature of various different agents within a GCM and between GCMs.

Here, adjusted radiative forcing is used, referred to as the effective radiative forcing (ERF) in IPCC AR5 (Myhre et al., 2013). This includes both stratospheric and tropospheric/land-surface adjustments (Gregory and Webb, 2008; Andrews et al., 2012). While the traditional radiative forcing concept only allowed for the adjustment of stratospheric temperatures to radiative equilibrium, ERF allows for tropospheric temperatures, water vapour and clouds to adjust. These rapid adjustments occur on a timescale of days to months and are separable from the more

unresponsive surface temperature changes. Typically, differences between radiative forcing and ERF are only noticeable for aerosols, because of the number of associated tropospheric rapid adjustments, including aerosol-cloud microphysical interactions and changes to the cloud lifetime (see Section 1.4).

2.2.1 Calculating forcing and feedback

Change in net downward radiation is defined at the TOA here, using one of two common techniques for calculating ERF. One can either fix SST and allow other components of the climate system to vary until a steady state is reached (Hansen et al., 2005), or one can use a linearised global energy budget technique (initially proposed by Gregory et al. (2004)) to determine forcing and feedback in response to an instantaneous perturbation. Here, the linearised global energy budget approach (described in Forster and Taylor (2006) and updated in Forster et al. (2013)) is used to derive forcing time series in CMIP5 models. The net TOA energy imbalance (N), following an initial radiative perturbation (F), is dependent on global mean surface air temperature change (ΔT), such that

$$N = F - \alpha \Delta T, \tag{2.3}$$

where α ($\mathrm{Wm^{-2}\ K^{-1}}$) is the climate feedback parameter. N and ΔT global annual means are taken directly from the abrupt4×CO_2 experiments and anomalies are taken with respect to the mean global annual mean of each model's preindustrial control simulation. In Forster et al. (2013) for a given CMIP5 model any drift – calculated as a linear trend – in the corresponding segment of the preindustrial control is subtracted from the N and ΔT values in the abrupt4×CO_2 experiment. Ignoring this extra step (the times at which perturbed runs branch from the control are not well documented in CMIP5 data files), near-identical forcing and feedback values to those in Forster et al. (2013) are calculated (Table 2.4).

Using this idealised CO_2-only simulation, values of N are regressed against ΔT to

give an estimate of F, as the intercept, and α, as the slope (Figure 2.2). Estimates for the 23 CMIP5 models (Table 2.1) are shown in Table 2.4. Using global annual means of N and ΔT from the historical all forcings simulations – the mean across ensemble members is taken for models where multiple initial condition runs were made – and diagnosed α values, equation (2.3) is rearranged to give

$$F = N + \alpha \Delta T, \tag{2.4}$$

with F calculated over the period 1905-2004 (Figure 2.3). This approach relies on the key assumption that α is time invariant (see Section 5.2 for a discussion about this assumption) and independent of the forcing agent. From these calculations one can also readily determine the ECS ($\Delta T_{2\times}$), the change in the global mean surface air temperature following a doubling of CO_2, in each model. It is simply the temperature change when the energy balance has reached equilibrium, i.e. $N = 0$. Here, $\Delta T_{2\times} = \frac{F}{2\alpha}$, the 2 in the denominator adjusting for the fact that CO_2 has been quadrupled rather than doubled. ECS is another useful metric to compare climate models (Table 2.4).

2.2.2 Calculating longwave and shortwave components

The same linearised global energy budget approach can be used for calculating the LW and SW components of the net downward radiative forcing, but instead using LW TOA energy imbalance (N_{LW}) and SW TOA energy imbalance (N_{SW}) in equation (2.3) to calculate the LW climate feedback parameter (α_{LW}) and the SW climate feedback parameter (α_{SW}) respectively (Figure 2.2). These parameter values can then be applied to equation (2.4) to calculate F_{LW} and F_{SW}, giving LW and SW time series respectively (Figure 2.3), the sum of which is constrained to equal F.

	Forcing (F) ($\mathrm{Wm^{-2}}$)	Feedback (α) ($\mathrm{Wm^{-2}\ K^{-1}}$)	ECS ($\Delta T_{2\times}$) (K)
ACCESS1-0	2.94	0.76	3.85
bcc-csm1-1	3.38	1.22	2.77
CanESM2	3.81	1.05	3.63
CCSM4	3.65	1.27	2.88
CNRM-CM5	3.68	1.15	3.20
CSIRO-Mk3-6-0	2.54	0.60	4.22
FGOALS-s2	3.73	0.90	4.15
GFDL-CM3	3.04	0.78	3.90
GFDL-ESM2G	3.35	1.44	2.33
GFDL-ESM2M	3.33	1.36	2.46
GISS-E2-H	3.77	1.60	2.36
GISS-E2-R	3.67	1.72	2.14
HadGEM2-ES	2.99	0.67	4.45
inmcm4	3.05	1.54	1.99
IPSL-CM5A-LR	3.21	0.79	4.06
IPSL-CM5A-MR	3.33	0.81	4.10
IPSL-CM5B-LR	2.64	1.01	2.61
MIROC5	4.38	1.65	2.66
MIROC-ESM	4.31	0.93	4.66
MPI-ESM-LR	4.19	1.16	3.61
MPI-ESM-P	4.39	1.28	3.44
MRI-CGCM3	3.31	1.29	2.57
NorESM1-M	3.13	1.14	2.74
Multi-model mean	**3.47**	**1.13**	**3.24**
5–95% uncertainty	**± 0.86**	**± 0.53**	**± 1.34**

Table 2.4: Estimates of F, α and ECS from 23 CMIP5 models (Table 2.1), using the abrupt4×CO_2 experiments. Estimates of F and ECS have been scaled to 2×CO_2 conditions by halving (ECS is by definition the change in the global mean surface air temperature following a doubling of CO_2). For the multi-model mean the error (5–95% uncertainty range) is calculated from the range of best estimates across the 23 models.

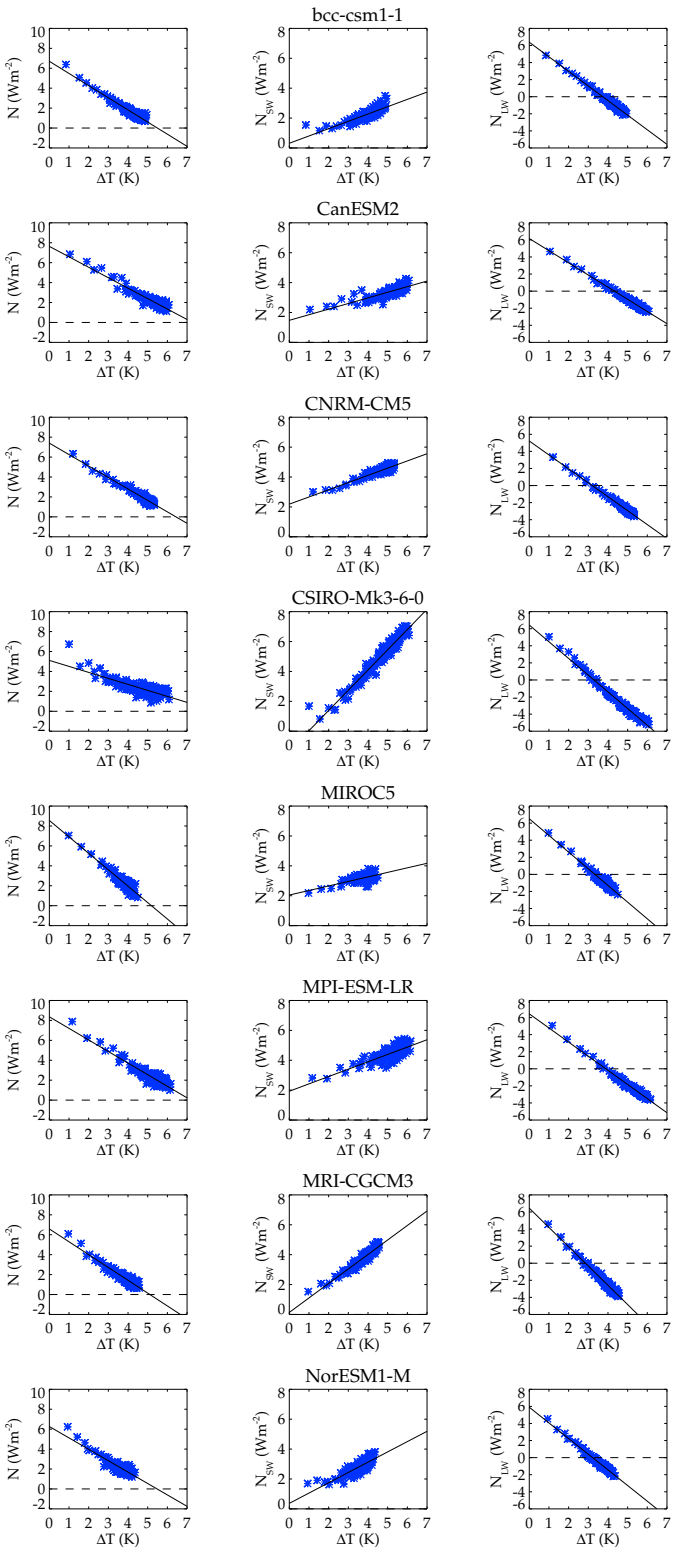

Figure 2.2: Relationships between net TOA energy imbalance (N) and global mean surface air temperature (ΔT) (left column), LW TOA energy imbalance (N_{LW}) and global mean surface air temperature (ΔT) (middle column) and SW TOA energy imbalance (N_{SW}) and global mean surface air temperature (ΔT) (right column), after an instantaneous quadrupling of CO_2 in the idealised abrupt4×CO_2 experiments. Data points are global annual means. Lines represent ordinary least squares regression fits to 150 years of data. The intercept at $\Delta T = 0$ (left column) is equal to the radiative forcing (F) (in Table 2.4 this has been scaled to 2×CO_2 conditions by halving) and the gradient is equal to the climate feedback parameter (α). Anomalies are with respect to the mean global annual mean of each models preindustrial control simulation. A subset of 8 of the 23 CMIP5 models is shown.

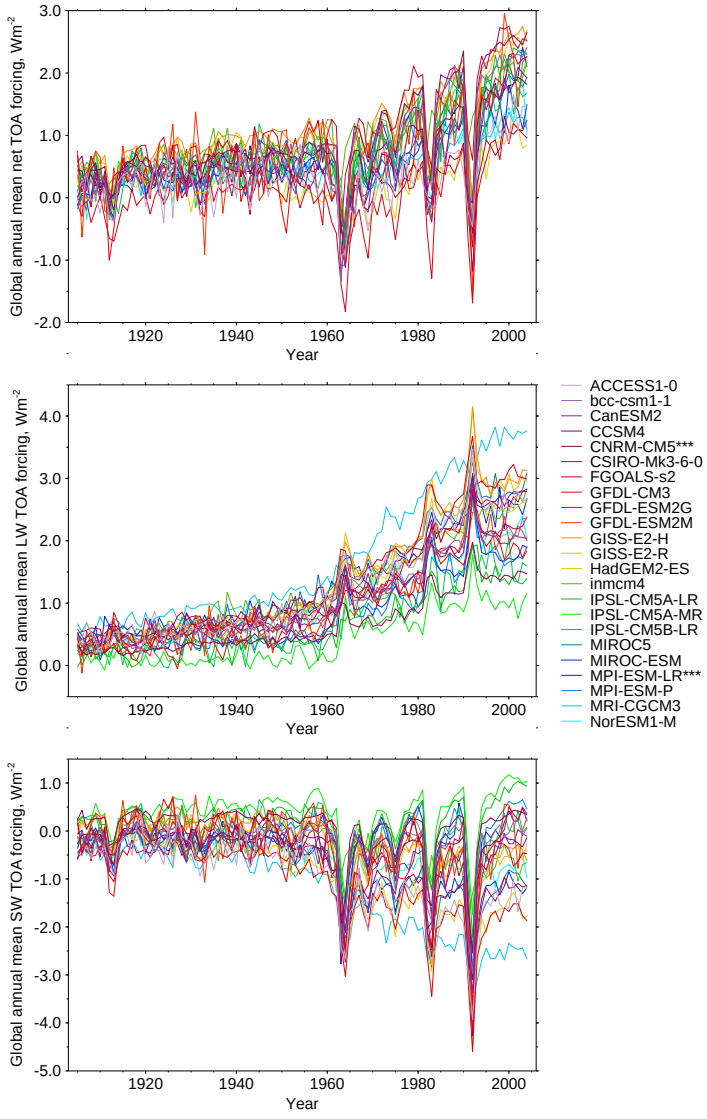

Figure 2.3: Time series of net (top), LW (middle) and SW (bottom) global annual mean TOA radiative forcing for the twentieth century (1905-2004). Global annual mean forcing is calculated using N and ΔT values from the historical all forcings simulations and α values diagnosed from the abrupt4×CO_2 experiments (Table 2.4), as in equation (2.4). All 23 CMIP5 models are shown. The two CMIP5 models considered further in Section 2.3.1 (Figure 2.5) have asterisks next to them.

2.2.3 Radiative forcing in CMIP3

To comment on the evolution of forcing, feedback and climate sensitivity in the latest climate models, a comparison is made with models forming part of CMIP3. TOA data for F, F_{LW} and F_{SW} (all time series), as well as α were available (data were kindly provided by Mat Collins), but, unlike in CMIP5, these were calculated from just one of the ensemble members (run 1) from each model considered. This may increase the uncertainty in the calculated forcing time series. Corresponding precipitation and temperature output was also available. Data were available for 1905-1999, so finishing five years earlier than the CMIP5 simulations. The 17 CMIP3 models are summarised in Table 2.2.

The most obvious difference between the intercomparison projects is that CMIP5 considered a greater number of forcing agents. For example, nearly all CMIP5 models contain volcanic forcing, whereas it is included in only 9 of the 17 CMIP3 models analysed here (Forster and Taylor, 2006). Time-evolving solar forcing is also used in most CMIP5 models, an improvement on CMIP3, while many have added interactive ocean and land carbon cycles, although the complexity of these representations varies considerably between models (Sillmann et al., 2013). Multiple CMIP5 models now also include the first indirect effect of aerosols, while several even include the second aerosol indirect effect (Wilcox et al., 2013) (see Section 1.4). Finally, CMIP5 models generally have higher horizontal and vertical resolution than CMIP3 models, reflecting increases in computational power (Taylor et al., 2012). For a more detailed discussion comparing CMIP3 and CMIP5 see Section 1.5.3.

2.3 Defining the offset

We expect sulphate-driven cooling to lead to a greater reduction in precipitation than the increase in precipitation resulting from GHG-driven warming for a given

radiative forcing. The ΔT dependent precipitation response is independent of the climate change mechanism (see Section 2.1.1), but the direct effect of changing GHG concentrations is significant and opposes the temperature dependent precipitation change, muting the total precipitation response. However, the direct effect of sulphate aerosols is negligible, therefore the total precipitation response is greater in magnitude for a unit radiative forcing and subsequent change in temperature (see Section 2.1.2).

One can take a simplified view of twentieth-century forcing by considering just the GHG and aerosol components, the two forcing agents that caused the largest, albeit opposing, radiative forcings over this period (Myhre et al., 2013). Here, GHG forcing is approximated as increasing linearly over the twentieth century and aerosol forcing is approximated as a step change, from zero to a given negative forcing, in 1960. Because the radiative forcing by GHGs and aerosols occur mostly in the LW and SW spectrums respectively, this approximation can be verified in Figure 2.3. Although increasing monotonically, the global mean TOA LW forcing appears to accelerate around 1960. Otherwise, a linear increase would seem a fair approximation. Ignoring the strong negative spikes causes by explosive, stratosphere-injecting volcanic eruptions, most models show a shift towards more negative global mean TOA SW forcing around 1960, although there is clearly a wide spread in the magnitude of this shift. Again, the approximation suggested would seem to be reasonable. It does, however, neglect the period of decreased aerosol/SW forcing towards the end of the twentieth century when numerous clean air policies came into effect (known as global brightening) (Wild, 2012). But this does not stand out as an obvious feature. Furthermore, the influence of black carbon aerosols is ignored for now. But the potential implications of black carbon forcing on the tropospheric energy budget and global mean precipitation change are discussed in more detail later.

Given these forcing approximations a schematic of the expected global mean precipitation-temperature relationship over the twentieth century is shown in Fig-

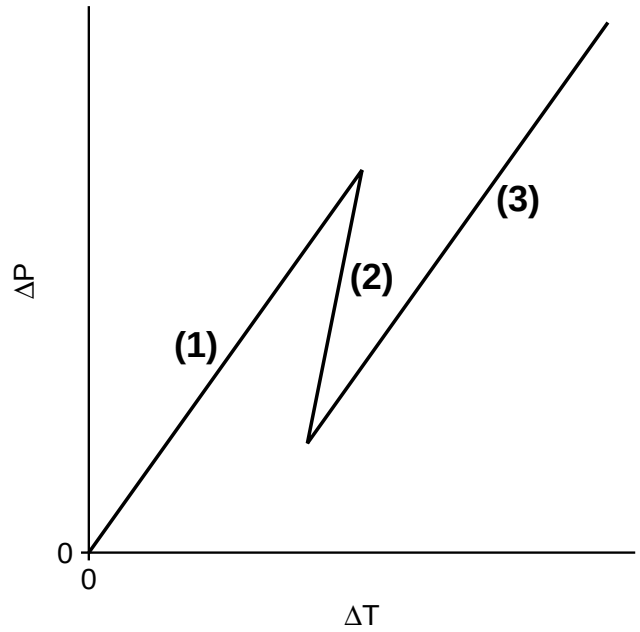

Figure 2.4: Schematic of the expected global mean precipitation-temperature relationship over the twentieth century, given approximate (idealised) GHG and aerosol forcing time series. Lines are numbered to represent the three prominent responses. See text for details.

ure 2.4, represented by the three lines (1), (2) and (3). Starting at the beginning of the twentieth century at the origin (with changes in precipitation and temperature relative to the preindustrial period), precipitation increases linearly with temperature as GHG forcing increases (1). Then, following a step increase in sulphate aerosol concentrations around the mid twentieth century, temperature decreases. But because the total precipitation is more sensitive to temperature under aerosol forcing (because of the negligible direct effect), there is a shift away from (1) and line (2) is traced. Thereafter, aerosol forcing is held constant, but GHG forcing continues to increase in time at the same rate as before. As a result, precipitation again increases linearly with temperature and line (3) is parallel to line (1), but "offset" because of the effect of the step change in aerosol. This forms the basis of our definition of the precipitation offset – two GHG-driven straight lines offset by mid-twentieth-century aerosol-driven cooling – used throughout this thesis, although it is altered somewhat (see Section 2.3.1.2) to reflect the far from idealised real-world response.

2.3.1 The offset in CMIP5 models

Only satellite-based precipitation datasets offer a picture of global mean precipitation (see Section 1.2.2) and the earliest records only began in the 1970s, after the rapid increase in aerosol forcing. Therefore, this idealised offset approach cannot be verified with real-world observations. But to what extent, if at all, does the global mean precipitation-temperature relationship in climate models show this response? Using five-year global mean precipitation and temperature in the twentieth century historical all forcings simulations, with anomalies taken with respect to the mean of the preindustrial control simulations, one sees evidence for this offset (Figure 2.5). Five-year means are used to improve the signal to noise ratio, as well as to avoid autocorrelation, which becomes a problem when using annual means.

Across the ensemble of 23 CMIP5 models the offset is most obvious when dividing the five-year means into a pre-1960 group and a post-1960 group (Figure 2.5 illustrates this point with two CMIP5 models). This is consistent with the marked increase in forcing evident in both the global annual mean real-world aerosol forcing time series (Figure 1.1) and global annual mean model simulated SW forcing time series (Figure 2.3). But interestingly the magnitude of this offset appears to vary considerably across the models, with two of the more extreme examples shown by the MPI-ESM-LR and CNRM-CM5 models in Figure 2.5. CMIP5 models, unlike CMIP3 models, all use the same aerosol emission inventory (Allen et al., 2013), that of Lamarque et al. (2010). But CMIP5 models still differ greatly in their radiative transfer schemes, microphysical parametrisations of aerosol-cloud interactions and whether they include interactive prognostic aerosols and chemistry (see Sections 1.4 and 1.5). Hence, the atmospheric aerosol loading, as well as radiative properties, are different and the range of aerosol forcing across the CMIP5 models is considerable.

The MPI-ESM-LR and CNRM-CM5 models have one of the smallest and one of the largest increases in mid-twentieth-century SW forcing respectively, across

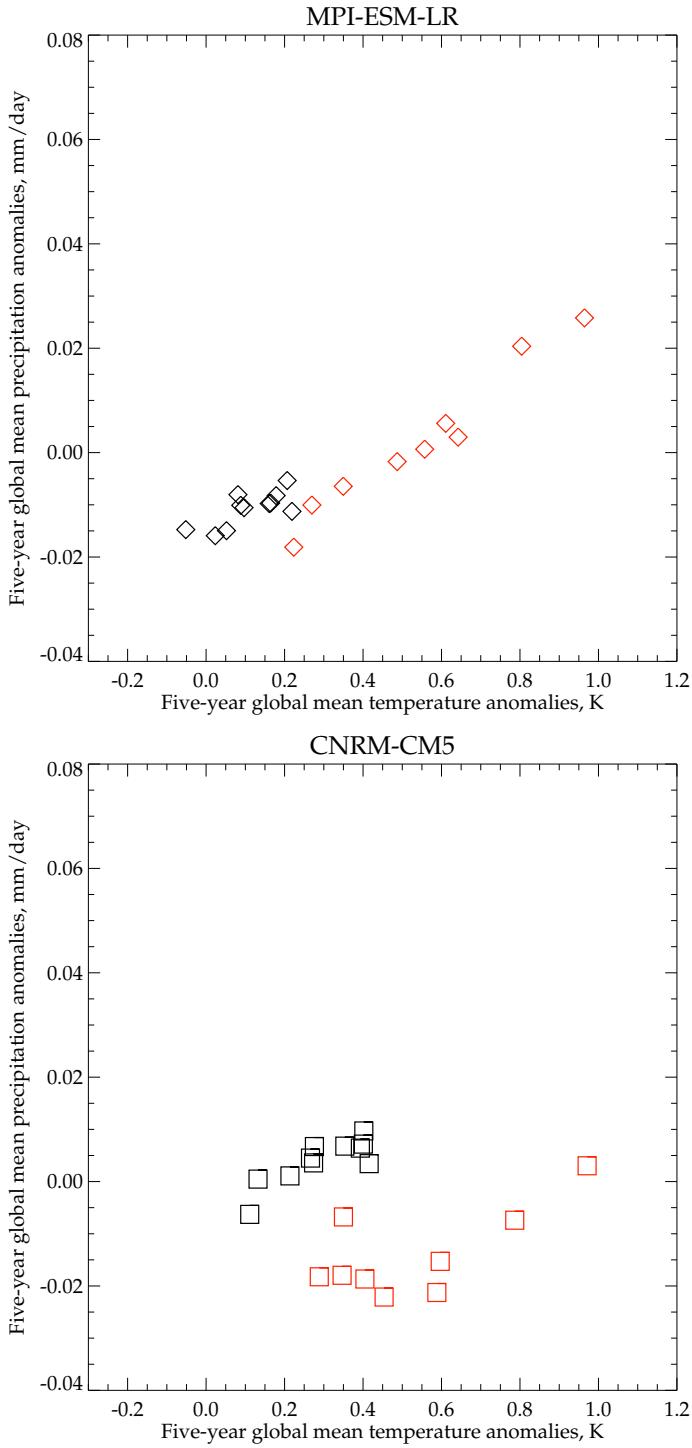

Figure 2.5: Five-year global mean precipitation-temperature relationship for the historical all forcings experiment with the MPI-ESM-LR model (top) and the CNRM-CM5 model (bottom) for 1905-2004. The mean across all ensemble members is taken for each model (Table 2.1). Anomalies are given relative to the mean of each model's preindustrial control simulation. Black; five-year means 1905-1959. Red; five-year means 1960-2004.

the whole CMIP5 ensemble (Figure 2.3). This seems to be reflected in the size of the offset in Figure 2.5. To investigate the extent to which this offset-forcing relationship holds across the whole CMIP5 ensemble, methods for diagnosing the magnitude of the increase in mid-twentieth-century aerosol forcing (referred to as the strength of aerosol forcing) in models and the magnitude of the precipitation offset in models are needed.

2.3.1.1 Calculating the strength of aerosol forcing

The strength of global mean TOA aerosol forcing, F_{AERO}, is calculated as

$$F_{AERO} = \overline{F_{SW_{60-04}}} - \overline{F_{SW_{05-59}}} \tag{2.5}$$

where $\overline{F_{SW_{60-04}}}$ is the 1960-2004 mean global mean TOA SW forcing and $\overline{F_{SW_{05-59}}}$ is the 1905-1959 mean global mean TOA SW forcing. To calculate the strength of aerosol forcing 5–95% uncertainty range, the variance of the difference of the two means, $Var(\overline{F_{SW_{60-04}}} - \overline{F_{SW_{05-59}}})$, is calculated as

$$Var(\overline{F_{SW_{60-04}}} - \overline{F_{SW_{05-59}}}) = Var(\overline{F_{SW_{60-04}}}) + Var(\overline{F_{SW_{05-59}}}) \tag{2.6}$$

where $Var(\overline{F_{SW_{60-04}}}) = \frac{s^2_{60-04}}{n_{60-04}}$ (s^2_{60-04} is the standard deviation of 1960-2004 five-year mean global mean TOA SW forcing anomalies and n_{60-04} the sample size) and $Var(\overline{F_{SW_{05-59}}}) = \frac{s^2_{05-59}}{n_{05-59}}$ (s^2_{05-59} is the standard deviation of 1905-1959 five-year mean global mean TOA SW forcing anomalies and n_{05-59} the sample size). Five-year mean forcing anomalies are used for consistency with other analyses.

Because the aerosol radiative forcing occurs largely at SW frequencies and other SW agents make a relatively minor contribution to overall SW forcing, the net aerosol and SW time series should be closely matched. The above ideas can be tested by exploring twentieth-century historical anthropogenic aerosol forcings only simulations that were run by a few modelling centres. Figure 2.6 shows forcing time series from the CSIRO-Mk3-6-0 model. The global mean TOA net

forcing and global mean TOA SW forcing time series (from the anthropogenic aerosol forcings only experiment) are closely matched, confirming that forcing at SW frequencies accounts for the vast majority of total aerosol forcing. Some changes in LW radiation are expected through rapid adjustments and aerosol-cloud interactions (Boucher et al., 2013).

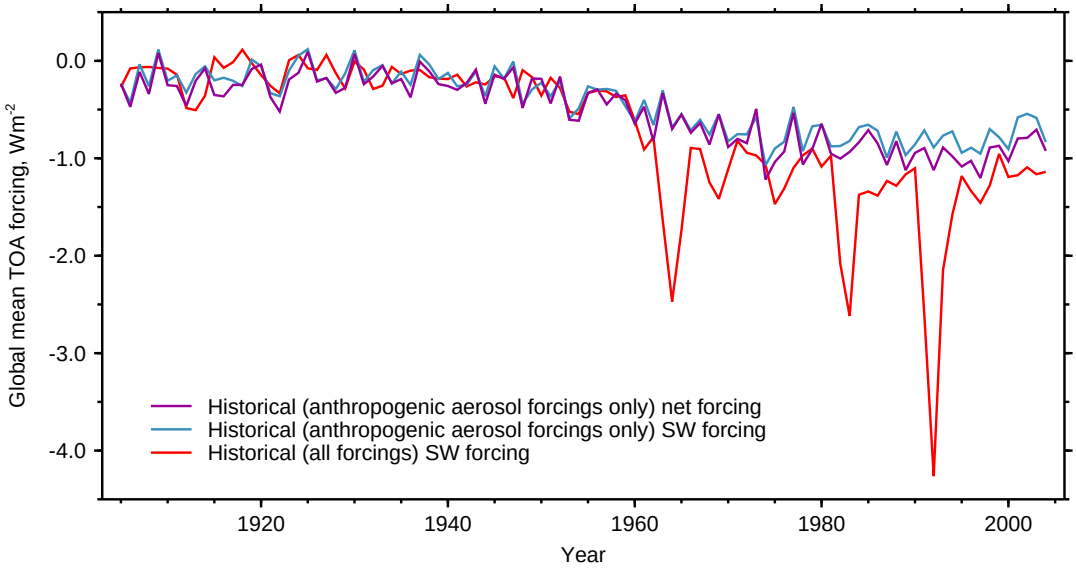

Figure 2.6: Time series of global mean TOA net forcing (purple) and global mean TOA SW forcing (blue) from the historical anthropogenic aerosol forcings only experiment and of global mean TOA SW forcing (red) from the historical all forcings experiment for the CSIRO-Mk3-6-0 model for 1905-2004. Forcing is given relative to a pre-industrial control simulation.

Further, the global mean TOA SW forcing time series are closely matched in both the historical anthropogenic aerosol forcings only and historical all forcings experiments (Figure 2.6). This indicates that aerosol SW forcing accounts for much of the total SW forcing in these models. In the absence of further idealised aerosol simulations it is difficult to separate the contributions from sulphate aerosols and black carbon aerosols. But black carbon aerosols absorb strongly in the atmosphere and can cause a positive SW forcing at the TOA but a negative SW forcing at the surface (Ramanathan and Carmichael, 2008). This is in contrast to sulphate aerosols where TOA and surface forcing should be of the same sign and roughly the same magnitude (Andrews et al., 2010).

To test for significant black carbon forcing in such simulations one can consider the ratio of TOA SW forcing to surface SW forcing in time. Surface forcing is

diagnosed using the method outlined in Section 2.2.1, but using model simulated surface radiative fluxes instead of TOA radiative fluxes. This follows the method of Andrews et al. (2009). The presence of black carbon would show as a shift in this ratio away from unity. There is no real evidence for this though (not shown) in the few models with dedicated anthropogenic aerosol only simulations, suggesting that the aerosol mix is dominated by sulphate aerosols. So, even though black carbon, other SW agents and the simple method for calculating the forcing will cause some uncertainty in the estimates, these values are expected to be representative of the magnitude of the mid twentieth century increase in sulphate aerosol forcing.

2.3.1.2 Calculating the precipitation offset

Revisiting Figure 2.4, the size of the global mean precipitation offset can be thought of as the vertical distance between the two parallel GHG-driven lines, (1) and (3). Therefore, the offset is simply equal to the difference in the intercepts of line (3) and line (1). But problems with this difference in intercepts approach become apparent when fitting a simple linear regression model to the actual data to calculate the precipitation offset.

Fitting separate regression lines to the pre-1960 data and the post-1960 data (the groupings used in Figure 2.5) the pre-1960 data are modelled as

$$P_{1i} = \beta_1 + \gamma_1 T_{1i} + \epsilon_{1i} \tag{2.7}$$

where P_1 are five-year mean global mean precipitation anomalies for 1905-1959, T_1 are five-year mean global mean temperature anomalies for 1905-1959, γ_1 is the trend from the linear regression fit to 1905-1959 global mean precipitation-global mean temperature, β_1 is the intercept and the residuals ϵ_{1i} are assumed to be identically and independently distributed with zero mean. For the post-1960 data

$$P_{3i} = \beta_3 + \gamma_3 T_{3i} + \epsilon_{3i} \tag{2.8}$$

where data and regression coefficients are as in (2.7), but for 1960-2004.

In the situation that $\gamma_1 = \gamma_3$ (i.e. parallel lines), the vertical distance between the two lines is always $\beta_3 - \beta_1$. However, if $\gamma_1 \neq \gamma_3$ the vertical distance between the two lines will vary with T. Therefore, the precipitation offset, if calculated as $\beta_3 - \beta_1$, will be dependent on the period to which anomalies are taken relative to. If confidence intervals of either regression function are wide, then the precipitation offset calculation will be highly uncertain. Across models the confidence intervals of the post-1960 regression function are narrower than the confidence intervals of the pre-1960 regression function. This is a consequence of the large increases in global mean temperature after the mid twentieth century increase in aerosol forcing, a consequence of the real world acceleration in GHG forcing. But increases in global mean temperature in the early twentieth century are less apparent and the global mean precipitation response less robust (reduced signal to noise); for comparison, in the abrupt4×CO_2 experiments shown in Figure 2.1 the global mean precipitation sensitivity to global mean temperature change, defined as k_T, is robust in all models, due to the very large ΔT increases (this is reflected in the small uncertainty ranges in Table 2.1). In Chapter 3, where regional land mean precipitation is considered and the precipitation offset is also calculated, a robust relationship with global mean temperature is only evident in the post-1960 period (in both CMIP5 models and observations).

To be consistent across analyses, a linear regression fit is made to 1960-2004 data and the (global mean or otherwise) precipitation offset (δ) is measured as the vertical distance between the regression line and the point representing the mean of the 1905-1959 data. This is equivalent to the intercept of the 1960-2004 linear regression line, with anomalies for the 1960-2004 period relative to the 1905-1959 mean. So

$$P_{\tilde{3}i} = \delta + \gamma T_{\tilde{3}i} + \epsilon_{\tilde{3}i} \tag{2.9}$$

where (here considering *global mean* precipitation) $P_{\tilde{3}i}$ are five-year mean global mean precipitation anomalies for the 1960-2004 period with respect to the 1905-

1959 mean global mean precipitation anomaly and $T_{\tilde{3}i}$ are five-year mean global mean temperature anomalies for the 1960-2004 period with respect to the 1905-1959 mean global mean temperature anomaly. The intercept term, δ, is the global mean precipitation offset, a measure of the precipitation response to aerosol forcing, and γ is the slope from the linear regression fit to 1960-2004 five-year means. The residuals $\epsilon_{\tilde{3}i}$ are assumed to be identically and independently distributed with zero mean.

To calculate 5–95% uncertainty ranges for δ, the variance of the conditional mean of the predictand, $s^2_{\overline{P_{\tilde{3}}}}$, given a particular predictor value $T_{\tilde{3}} = T^*$ is considered (Wilks, 2006)

$$s^2_{\overline{P_{\tilde{3}}}|T_{\tilde{3}}=T^*} = s^2_\epsilon \left[\frac{1}{n} + \frac{(T^* - \overline{T_{\tilde{3}}})^2}{\sum_{i=1}^{n} (T_{\tilde{3}i} - \overline{T_{\tilde{3}}})^2} \right] \tag{2.10}$$

where n is the sample size, $\overline{T_{\tilde{3}}}$ is the 1960-2004 mean global mean temperature anomaly and the square of the residual standard error, s^2_ϵ, is

$$s^2_\epsilon = \frac{1}{n+2} \sum_{i=1}^{n} \epsilon_{\tilde{3}i} \tag{2.11}$$

To estimate the variance of the intercept, $s^2_{\hat{\delta}}$, the case when $T^* = 0$ in equation (2.10) is considered, so

$$s^2_{\hat{\delta}} = s^2_\epsilon \left[\frac{1}{n} + \frac{\overline{T_{\tilde{3}}}^2}{\sum_{i=1}^{n} (T_{\tilde{3}i} - \overline{T_{\tilde{3}}})^2} \right]. \tag{2.12}$$

Therefore, the best estimate of the precipitation offset and the 5–95% uncertainty range is $\delta \pm 1.64 s_{\hat{\delta}}$.

2.4 The global mean precipitation response to aerosol forcing

2.4.1 CMIP5

There exists a remarkably strong correlation ($r = 0.80$, $p < 0.01$), where r is the Pearson correlation coefficient and p is the significance level, between the global mean precipitation offset and the strength of aerosol forcing across 23 CMIP5 models (Figure 2.7). The stronger the global TOA aerosol forcing in a model the greater the departure from an otherwise linear GHG-driven response and the greater (more negative) the global mean precipitation offset, as hypothesised in Section 2.3.

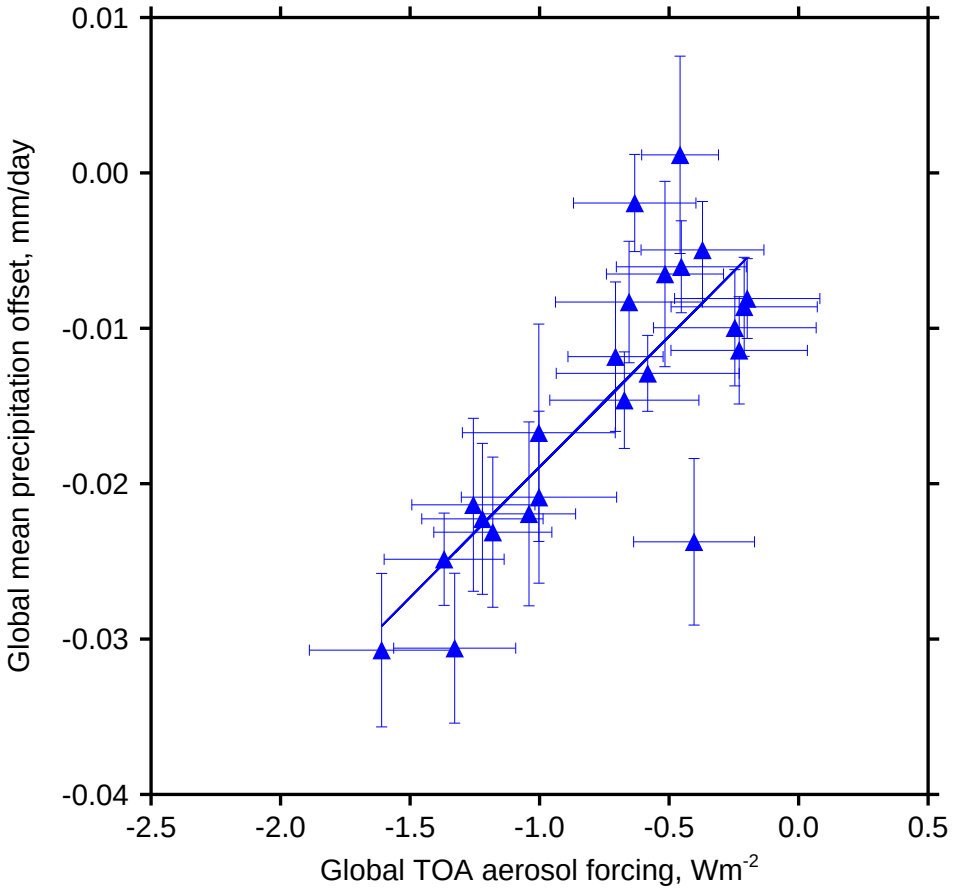

Figure 2.7: Global mean precipitation offset against global TOA aerosol forcing. Each data point represents one CMIP5 model, with 23 CMIP5 models in total (Table 2.1). Error bars represent the 5–95% uncertainty range (see Sections 2.3.1.1 and 2.3.1.2).

There is, however, one glaring outlier, the CNRM-CM5 model, which has a pre-

cipitation offset of -0.024 mm day^{-1} but an aerosol forcing of -0.40 Wm^{-2}. This could be because the precipitation offset is driven by strong black carbon aerosol forcing, through a significant direct effect on the tropospheric energy budget (see Section 2.1.2), and not just strong sulphate aerosol forcing. Because black carbon TOA forcing can be positive while surface forcing is negative, this might explain the anomaly.

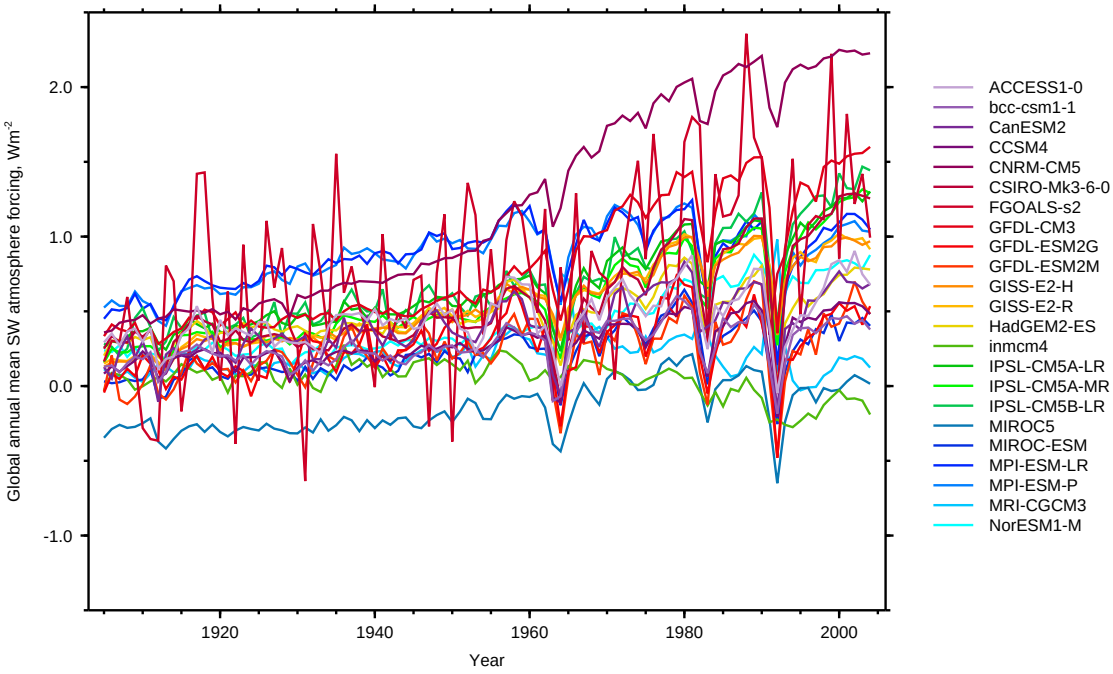

Figure 2.8: Time series of global annual mean SW atmosphere radiative forcing for the twentieth century (1905-2004). Atmosphere forcing is calculated as the difference between TOA and surface forcing (see text). All 23 CMIP5 models are shown.

Support for this idea comes from Figure 2.8. Here, global annual mean SW atmosphere radiative forcing is calculated as the difference between global annual mean SW TOA radiative forcing (Figure 2.3 (bottom panel)) and global annual mean SW surface radiative forcing, which is also calculated using the method outlined in Section 2.2.2, but using surface radiative fluxes rather than TOA radiative fluxes. The CNRM-CM5 model has, by far, the strongest atmosphere SW forcing in the second half of the twentieth century. This hints towards a significant black carbon aerosol contribution.

Because both increasing black carbon aerosol and increasing sulphate aerosol atmospheric concentrations lead to negative surface aerosol forcing and – albeit through different mechanisms – are both expected to generate a negative precip-

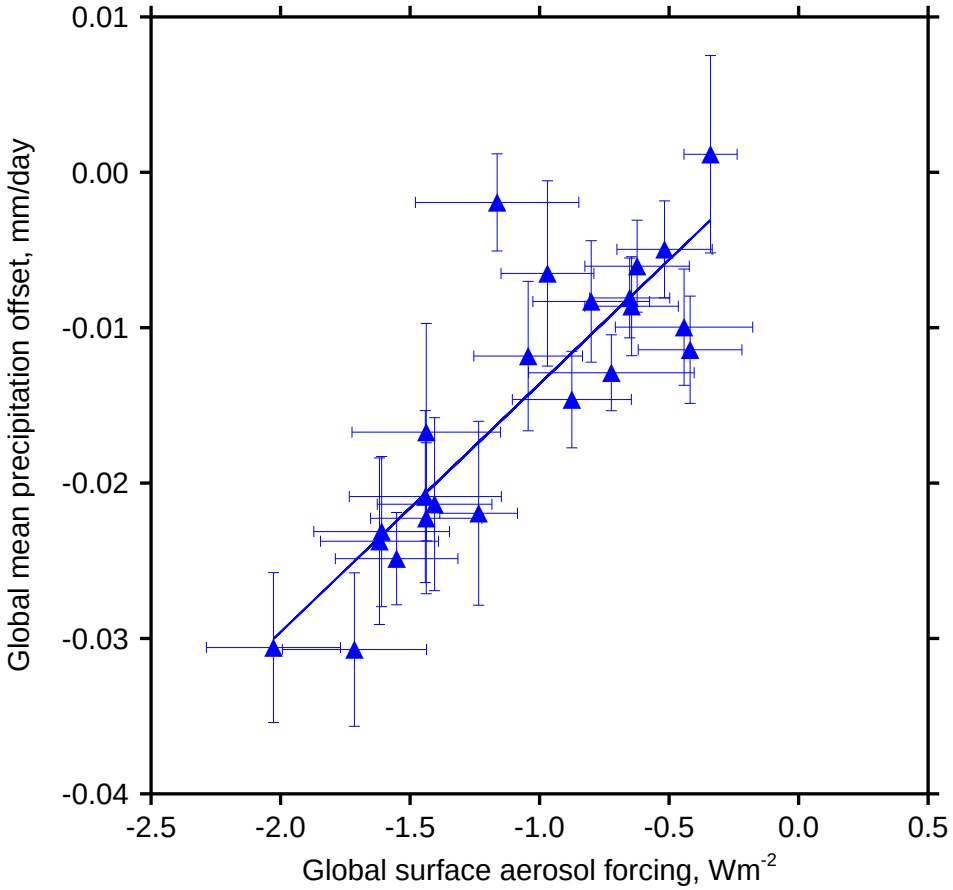

Figure 2.9: Global mean precipitation offset against global surface aerosol forcing. Each data point represents one CMIP5 model, with 23 CMIP5 models in total (Table 2.1). Error bars represent the 5–95% uncertainty range (see Sections 2.3.1.1 and 2.3.1.2).

itation offset in the mid twentieth century, the offset-forcing correlation is even stronger ($r = 0.86$, $p < 0.01$) when considering surface aerosol forcing (Figure 2.9) However, this is largely a consequence of the CNRM-CM5 model no longer being an outlier.

2.4.2 CMIP3

Also shown is the global mean precipitation offset against global TOA aerosol forcing for 17 CMIP3 models (Figure 2.10). A similar, albeit less robust ($r = 0.72$, $p < 0.01$), relationship is evident. Unfortunately, surface forcing data were not available for the CMIP3 models, so a similar comparison to that made above for CMIP5 is not possible. There are a few possible reasons for this lower correlation. Because precipitation and temperature data are only available for 1905-1999 for

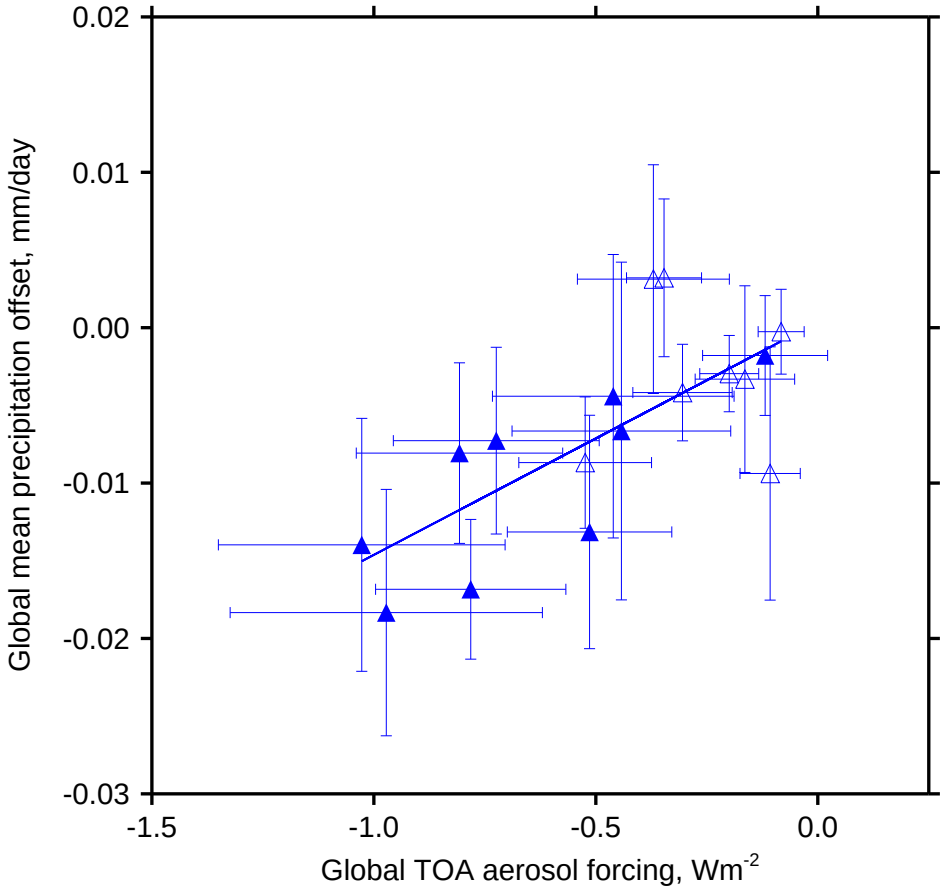

Figure 2.10: Global mean precipitation offset against global TOA aerosol forcing. Each data point represents one CMIP3 model, with 17 CMIP3 models in total. Filled points are models including a representation of volcanic forcing and unfilled points are models without volcanic forcing. Error bars represent the 5–95% uncertainty range (see Sections 2.3.1.1 and 2.3.1.2). Note that the axes ranges are different to those in Figures 2.7 and 2.9.

these CMIP3 simulations the sample size of the post-1960 period is reduced, increasing uncertainty in both variables. Only one of the ensemble members (run 1) from each model is included here, because of the availability of data. Another contributory factor could be the reduced range of global TOA aerosol forcing in CMIP3 models, relative to CMIP5 models. The CMIP5 models range from about -1.6 Wm^{-2} to -0.2 Wm^{-2} (Figure 2.7) and the CMIP3 models range from about -1.0 Wm^{-2} to -0.1 Wm^{-2} (Figure 2.10).

The influence of volcanic forcing may also play a role in the offset-forcing relationship. Optimal fingerprinting studies have detected the influence of volcanic forcing on observed global land mean precipitation (Gillett et al., 2004; Lambert et al., 2004) and model simulated global mean precipitation (Iles et al.,

2013). This is expected from an energetic constraint viewpoint. The three major stratosphere-injecting volcanic eruptions in the 1905-1999 period all happened post-1960. These are classed as the eruptions that caused an exceedance in global monthly mean stratospheric aerosol optical depth at 550 nm of 0.05 (see Sato et al. (1993) with updates available at http://data.giss.nasa.gov/modelforce/strataer/), giving Mount Agung (1963), El Chichón (1982) and Mount Pinatubo (1991). The 1913 eruption of Katmai was another notable event in the 1905-1999 period and only just falls outside this classification.

These eruptions in the post-1960 period could lead to a further reduction in global mean precipitation, because the total precipitation is also more sensitive to temperature under volcanic aerosol forcing than GHG forcing. However, these shifts are expected to be short-lived as volcanic forcing relaxes back to near-zero a couple of years after these explosive events. Because five-year means are considered in offset and forcing calculations, the overall influence of volcanic forcing on the offset-forcing relationship across models should be minimal. However, the strength of global TOA aerosol forcing does appear to be greater in CMIP3 models that include volcanic forcing – the global mean precipitation offset-global TOA aerosol forcing correlation also increases to $r = 0.77$ ($p < 0.05$) when considering just these 9 models – than in CMIP3 models that do not (Figure 2.10).

Beyond the difference in the range of the strength of global TOA aerosol forcing in CMIP3 and CMIP5, should the offset-aerosol relationship actually be the same between these intercomparison projects? There are many other differences between CMIP3 and CMIP5. This is the focus of the next section, where the evolution of forcing, feedback and climate sensitivity is considered.

2.5 Forcing and feedback

The perceived quality of a climate model is typically based on its ability to reproduce some aspect of the observed climate. Accurate simulation of observed

global temperature change over the twentieth century is frequently considered a prerequisite for an accurate representation of anthropogenic climate change and, in turn, for making reliable twenty-first-century projections.

2.5.1 Global mean temperature controlling factors

In Section 2.2.1 a linearised global energy budget approach was used to derive forcing time series. Using values of F and α, diagnosed by regressing N against ΔT (equation (2.3)) in the abrupt4×CO_2 experiments, the ECS, $\Delta T_{2\times}$, is given by $\Delta T_{2\times} = \frac{F}{2\alpha}$. The range in ECS will be smaller if there exists a correlation between F and α. This was noted by Andrews et al. (2012) when they found that the ECS range was much larger than the actual ECS range when holding F at the multi-model mean value (using a subset of 15 of the 23 CMIP5 models used here), implying a correlation between forcing and feedback. Here, a significant correlation is found between F and α ($r = 0.45$, $p < 0.05$).

The ECS, by definition, is the temperature change when the energy balance has reached equilibrium (following a doubling of CO_2). In this new steady state $N = 0$. But the real, transient climate is never in a steady state. Therefore, rearranging equation (2.3), $\Delta T = \frac{F-N}{\alpha}$. Here, F represents any radiative perturbation due to climate forcing, rather than the initial radiative perturbation following a quadrupling of CO_2 (as used in the context above). This means that N, which can also be thought of as the net heat flux into the climate system also has an influence. Heat is stored in the oceans, so this ocean heat uptake parameter moderates the rate of time-dependent climate change.

Rewriting ECS as $\Delta T_{2\times} = \frac{F_{2\times}}{\alpha}$, where $F_{2\times}$ is (specifically) the forcing due to a doubling of the CO_2 concentration, and using $\alpha = \frac{F_{2\times}}{\Delta T_{2\times}}$, equation (2.3) can be written as

$$F = \frac{\Delta T F_{2\times}}{\Delta T_{2\times}} + N. \tag{2.13}$$

$F_{2\times}$ is reasonably consistent across CMIP5 models ($F_{2\times} = 3.47 \pm 0.86$ Wm^{-2} from

column 2 of Table 2.4), as well as CMIP3 models (Forster and Taylor, 2006), while a small range in ΔT is a consequence of models accurately simulating the observed temperature trend. ECS ($\Delta T_{2\times}$), forcing (F) and ocean heat uptake (N) are all uncertain. But no correlation between climate sensitivity and ocean heat uptake across models should be expected because these quantities are diagnostic rather than tunable (Knutti, 2008). However, uncertainties in ocean heat uptake tend to be small compared with uncertainties in the applied forcing over the twentieth century and in the climate response to this forcing (Dufresne and Bony, 2008; Anderson et al., 2010). Fixing both the numerator of the fraction ($\Delta T F_{2\times}$) and N in equation (2.13) one can fit a theoretical, inverse relationship between net forcing (F) and ECS ($\Delta T_{2\times}$) (e.g. Kiehl (2007)).

Knutti (2008) found evidence of this inverse relationship between forcing and ECS in CMIP3, in agreement with the work of Kiehl (2007) on an older set of models. Forster et al. (2013) found evidence that this correlation persists in CMIP5 models, particularly when considering a subset of models that best match the observed twentieth-century global temperature trends. Similarly, taking $\Delta T = \frac{F-N}{\alpha}$ and assuming near-constant N, one could expect a quasi-linear relationship between forcing F and feedback α across models. There is found to be a significant correlation between α and F in both CMIP3 ($r = 0.69$, $p < 0.01$; note that CMIP3 correlations in this section exclude the BCCR-BCM2.0 model, because it has an anomalously large climate feedback parameter – it was also not included in the Forster and Taylor (2006) analysis) and CMIP5 ($r = 0.57$, $p < 0.01$) (Figure 2.11). This means that models with smaller feedbacks/higher sensitivities tend to have smaller forcings. The result is a narrower range in historical temperature trends than there would be with no correlation.

But what about the impact of ocean heat uptake? Previously it was considered as near-constant and noted that uncertainties in ocean heat uptake tend to be small compared to uncertainties in forcing and feedback (Dufresne and Bony, 2008). In a time-dependent state the ocean heat uptake N can be represented approximately

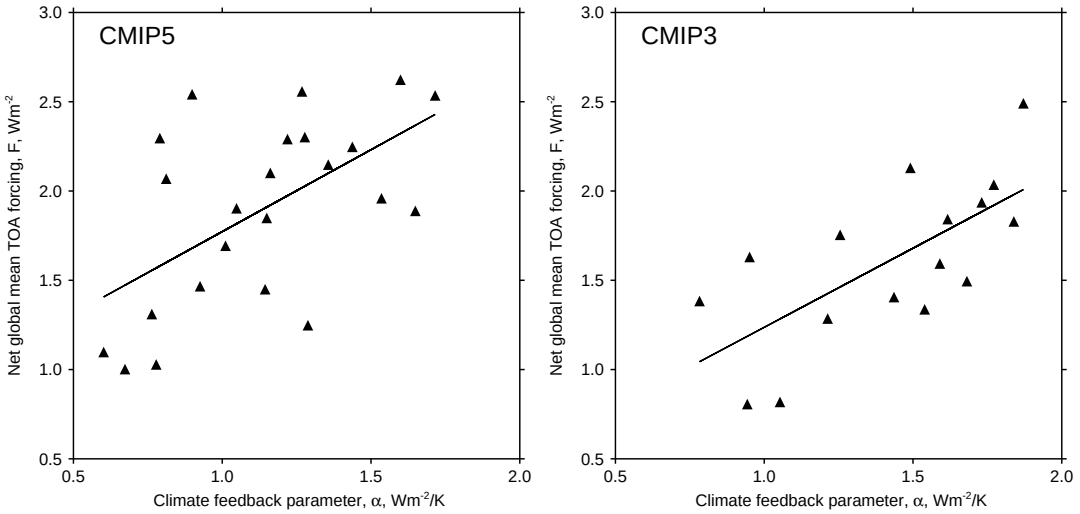

Figure 2.11: Scatterplots of end of twentieth-century net global mean TOA forcing, F, against the climate feedback parameter, α, in CMIP5 (left) and CMIP3 (right). Forcing is for the year 2002 (2000-2004 average) and 1997 (1995-1999 average) for CMIP5 and CMIP3 respectively, due to availability of data.

by $N = \kappa\Delta T$, where κ is the ocean heat uptake efficiency. Therefore, following Gregory and Forster (2008), equation (2.3) can be written as $\Delta T = \frac{F}{\rho}$, where the climate resistance $\rho = \alpha + \kappa$. Using a similar set of CMIP5 models to those used here Forster et al. (2013) showed that $\frac{F}{\rho}$ was strongly correlated with 2003 (2001-2005 average) temperature change. Their estimates of κ were made from idealised 1% per year CO_2 increase simulations (κ values, and therefore ρ values, are available for 22 of the 23 CMIP5 models used here). κ is not correlated with F ($r = 0.09$, not significant), but there is some evidence of a correlation between ρ and F ($r = 0.49$, $p < 0.05$). For comparison, this is smaller than the α and F correlation ($r = 0.57$, $p < 0.01$), as expected. So, while there seems to be no tendency for ocean heat uptake efficiency to cause a convergence in CMIP5 historical temperature trends, there is some evidence of a trade off between forcing and feedback in CMIP5 models (i.e. models with larger end of twentieth-century forcing tend to have larger climate feedback parameters/smaller sensitivities).

2.5.2 Contrasting CMIP3 and CMIP5 controlling factors

That there is still evidence of such a trade off is interesting. Both Kiehl (2007) and Knutti (2008) suggested that a small or large net forcing is usually the result

of strong or weak negative aerosol forcing respectively. Indeed, in CMIP3 the correlation between 1997 (1995-1999 average) net forcing and 1997 SW forcing is $r = 0.72$ ($p < 0.01$). Aerosol/SW forcing has remained uncertain through several generations of GCMs (see Section 1.4), hence the idea that choices over parameterisations could be made based on the feedback/sensitivity of a model to best replicate the observed temperature. Knutti (2008) suggested that the incorporation of more detailed aerosol effects between CMIP3 and CMIP5 could lead to a mismatch between simulated and observed warming over the twentieth century. This was based on the assumption that a more detailed consideration of aerosol effects would lead to less control over aerosol forcing. The CMIP5 multi-model mean 2002 SW forcing is -0.41 ± 1.45 Wm^{-2}, compared to the CMIP3 multi-model mean 1997 SW forcing of -0.56 ± 0.67 Wm^{-2}. Perhaps aerosol-cloud interactions have become so complex that the aerosol forcing is no longer predictable or adjustable (Shindell et al., 2013).

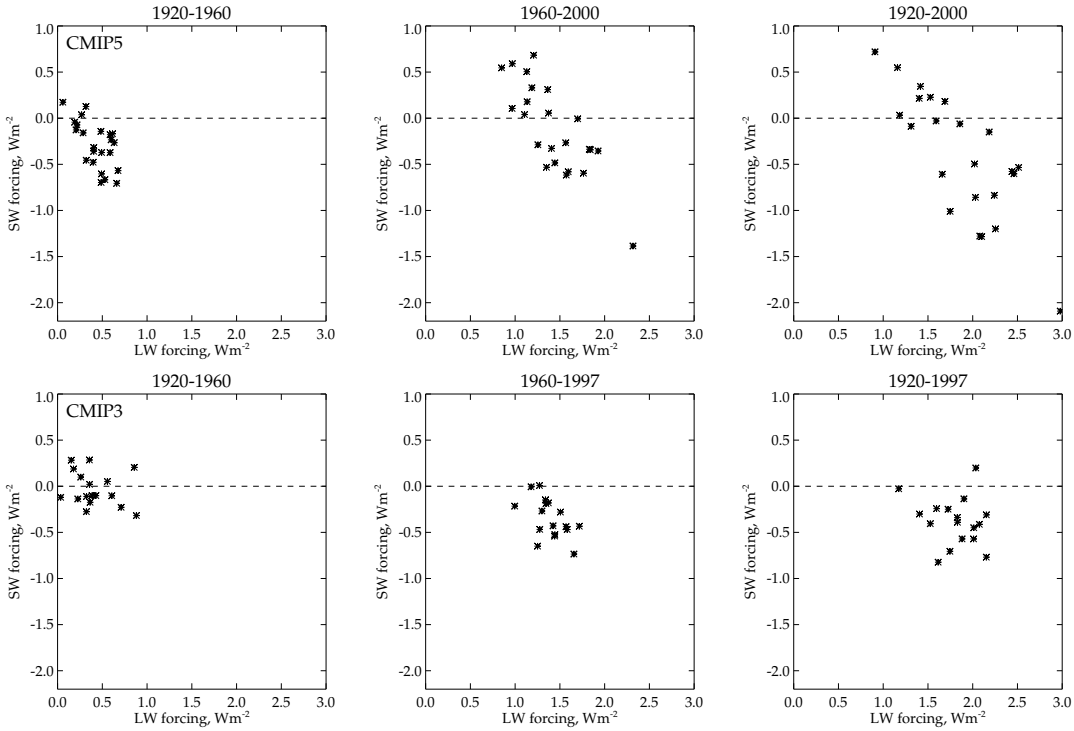

Figure 2.12: Changes in the LW and SW global mean TOA forcing from 23 CMIP5 models (top) and 16 CMIP3 models (bottom). These forcing changes are differences between two 5-year averages centred on the start and end years given. Because data are only available until 1999 for the CMIP3 models, the latest possible year, 1997, is used instead of the year 2000.

With little change in the spread of α between intercomparison projects, why are

CMIP5 models still producing the observed historical temperature trend so well? The increased uncertainty in SW forcing must be compensated by an increased spread in LW forcing (Figure 2.12). The CMIP5 multi-model mean 2002 LW forcing is 2.31 $\pm$ 1.07 Wm^{-2}, compared to the CMIP3 multi-model mean 1997 LW forcing of 2.17 $\pm$ 0.53 Wm^{-2}. The 5–95% uncertainty range in CMIP5 is more than twice that in CMIP3. Additionally, 2002 SW forcing and 2002 LW forcing are strongly anti-correlated in CMIP5 ($r = -0.81$, $p < 0.01$), but 1997 SW forcing and 1997 LW forcing are not correlated in CMIP3 ($r = -0.28$, not significant).

It is difficult to know whether this anti-correlation is a natural consequence of increased model complexity in CMIP5, or a product of model tuning. Miller et al. (2014) noted that GHG forcing in the GISS-E2-R CMIP5 model is nearly 15% larger than GHG forcing in the GISS-ER CMIP3 model, despite being prescribed with very similar CO_2, CH_4 N_2O and CFC concentrations. The radiative transfer scheme has been updated between models, but the forcing associated with a doubling of CO_2 is nearly identical. Miller et al. (2014) suggested that the difference is due to the other GHGs or their spectral overlap with CO_2. Although the CMIP5 ensemble includes ESMs using an interactive carbon cycle, the historical twentieth-century simulations were performed using prescribed globally averaged CO_2 concentration, not CO_2 emissions. Therefore, this aspect of model development cannot be attributed to the additional spread.

In summary, there is strong evidence that feedback/sensitivity and net forcing are correlated in the CMIP3 ensemble of models, with the strength of aerosol forcing key to this correlation. In CMIP5 there is a suggestion of a persisting correlation between feedback/sensitivity and net forcing, but it is less robust. The mean and range of both feedback/sensitivity and net forcing is similar across CMIP3 and CMIP5, despite double the spread in CMIP5 SW forcing, relative to CMIP3. Opposing this is a twofold increase in the range of CMIP5 LW forcing, relative to CMIP3, which is correlated with CMIP5 SW forcing.

2.6 The physical basis of the aerosol offset

Combining the simple model for global mean precipitation change in Section 2.3 with the forcing-feedback relationships in CMIP3 and CMIP5 outlined in Section 2.5, this section explores the precipitation offset in more detail. Given these obvious differences in CMIP3 and CMIP5 should the global mean precipitation response (the offset) to a given aerosol forcing actually be similar in the model ensembles?

Taking a similar approach to Lambert and Allen (2009) equation (2.2) can be further decomposed

$$L\Delta P \simeq k_T \Delta T + \Delta R_{LW} + \Delta R_{SW}, \tag{2.14}$$

where ΔR_{LW} represents increases in net radiative cooling due to changes in tropospheric LW absorbing species concentration and ΔR_{SW} represents net increases in cooling due to changes in tropospheric SW absorbing species concentration. From Section 2.1.2 it is known that the direct effect of changing CO_2 concentrations, and more generally changing GHG concentrations, on $L\Delta P$ is significant. This opposes the temperature dependent precipitation change, muting the total precipitation response. Because GHG absorption occurs overwhelmingly at LW frequencies $\Delta R_{LW} < 0$. Black carbon can strongly absorb SW radiation, so following similar arguments it can be expected that $\Delta R_{SW} < 0$. But in the simplified model of the global mean precipitation-temperature relationship over the twentieth century presented in Section 2.3 only two forcings are considered; GHG forcing is approximated as increasing linearly over the twentieth century and sulphate aerosol forcing is approximated as a step change around the mid twentieth century. Because the direct effect of sulphate aerosol forcing is thought to be negligible (see below), here it is assumed that $\Delta R_{SW} \simeq 0$.

Writing $\Delta R_{LW} = \beta_{LW} F_{LW}$ where F_{LW} is the LW global mean TOA forcing due to changes in GHG concentrations and β_{LW} is a unitless constant, equation (2.14)

becomes

$$L\Delta P \simeq k_T \Delta T + \beta_{LW} F_{LW}. \tag{2.15}$$

For comparison, Thorpe and Andrews (2014) refer to the direct effect term (ΔR_A in equation (2.2); $\beta_{LW} F_{LW}$ in equation (2.15) above) as a precipitation adjustment to the atmosphere radiative forcing. The precipitation adjustment (ΔP_{adj}) is written as

$$\Delta P_{adj} = -F_{ATM} = -\sum_i R_i F^i_{TOA}, \tag{2.16}$$

where R_i are the ratios of atmosphere to TOA forcing for different forcing agents and F^i_{TOA} are TOA forcing time series for different forcing agents. The product of R and F_{TOA} for a given forcing agent gives its direct tropospheric radiative forcing. Values of R for different forcing agents were calculated by Andrews et al. (2010) using idealised simulations. For CO_2 $R = 0.8$ (R is calculated as 0.5 for other GHGs) and for sulphate aerosols $R = 0$.

Because only GHGs and sulphate aerosols are considered in the simple model here, β_{LW} in equation (2.15) is analogous to R for CO_2/other GHGs. Likewise, writing $\Delta R_{SW} = \beta_{SW} F_{SW}$, R for sulphate aerosols is analogous to β_{SW}, justifying the above approximation that ΔR_{SW} is equal to zero. Note that Thorpe and Andrews (2014) adopt the sign convention that these precipitation adjustments mute the overall precipitation response through a decrease in net radiative cooling, whereas here, the direct effect represents increases in net radiative cooling. Therefore, these coefficients, for a given forcing, will be of opposite sign.

Considering the relationship between ΔP and ΔT, as in Figure 2.4, by dividing through by L – the latent heat of vapourisation, which is roughly equal to 29 Wm^2 mm^{-1} day in equivalent precipitation units – gives

$$\Delta P \simeq \frac{k_T}{L} \Delta T + \frac{\beta_{LW}}{L} F_{LW}, \tag{2.17}$$

where k_T/L (mm day^{-1} K^{-1}) is the sensitivity of ΔP to ΔT. All terms are in units of mm day^{-1}.

2.6.1 Deriving an equation for the offset

Building on the above ideas, this section explores the most important governing factors controlling the magnitude of the global mean precipitation offset. This is approached in an approximate, idealised manner, considering the dominant twentieth-century climate forcings of GHGs and aerosols. All GCMs (CMIP3 and CMIP5) are taken to simulate identical end of twentieth-century global mean temperature change, which is of course an aim of all modelling centres (see Section 2.5). Experimenting with varying two or more components from GHG forcing, sulphate aerosol forcing and feedback/sensitivity, while fixing k_T and β_{LW}, the offset is always found to be a function of sulphate aerosol forcing, as long as the same twentieth century global mean temperature change is simulated. Why?

Considering just lines (1) and (3) in Figure 2.4, line (1) can be written as

$$\Delta P_1 \simeq \frac{k_T}{L}\Delta T_1 + \frac{\beta_{LW}}{L}F_{LW_1}, \tag{2.18}$$

and line (3) can be written as

$$\Delta P_3 \simeq \frac{k_T}{L}\Delta T_3 + \frac{\beta_{LW}}{L}F_{LW_3}. \tag{2.19}$$

Because of the linear increase in GHG forcing, lines (1) and (3) are parallel and the offset is just the difference in the intercepts (see Section 2.3.1.2), so the global mean precipitation offset (δ) becomes

$$\delta = \frac{\beta_{LW}}{L}F_{LW_3} - \frac{\beta_{LW}}{L}F_{LW_1}, \tag{2.20}$$

But because the net forcing after the step change in sulphate aerosol forcing is equal to the sum of (linearly increasing) GHG forcing (F_{LW_3}) and (constant) sulphate aerosol forcing (F_{SW_2}), which at some temperature ($\Delta T_1 = \Delta T_3 = \Delta T^*$) is equal to the forcing from just GHG forcing alone prior to the step change (F_{LW_1}), in this idealised situation one can write $F_{LW_3} + F_{SW_2} = F_{LW_1}$. Therefore,

substituting into equation (2.20) gives

$$\delta = \frac{\beta_{LW}}{L}(F_{LW_1} - F_{SW_2}) - \frac{\beta_{LW}}{L}F_{LW_1}, \tag{2.21}$$

or

$$\delta = -\frac{\beta_{LW}}{L}F_{SW_2}. \tag{2.22}$$

This shows that the precipitation offset is a function of the product of β_{LW} and F_{SW_2} (L is constant) in this idealised situation. It is expected that both β_{LW} and F_{SW_2} are negative in the real world and in GCMs, so that the precipitation offset is negative. But how constant is β_{LW}? It has been shown that there is a wide range in the strength of global mean TOA aerosol forcing (see Section 2.4), analogous to F_{SW_2} here. This alone explains much of the variance in the global mean precipitation offset. Some of the variance will be explained by forcings and other complexities not explored in the simple model considered here.

β_{LW} can be estimated by using ΔR_A calculated from the abrupt4×CO_2 experiments (Table 2.3) and by writing $\Delta R_A \simeq \beta_{LW} F_{LW}$. Even though other climate forcings, especially other GHGs, will contribute towards the exact value of β_{LW} in a model, it is expected to be close to the value from CO_2 forcing alone. Andrews et al. (2010) estimated values for CO_2 and other GHGs to be -0.8 and -0.5 respectively, justifying this expectation. F_{LW} is taken to be F in Table 2.4. This is again an approximation since most, rather than all, CO_2 forcing occurs in the LW. The multi-model mean β_{LW} is -0.66 ± 0.26.

But β_{LW} can also be calculated from a transient scenario by building a regression model for global mean precipitation, as in Lambert and Allen (2009). This uses equation (2.15) as the linear regression model and five-year mean GCM temperatures and precipitation, as well as LW forcing time series (Table 2.5 – results for a subset of 8 of the 23 CMIP5 models are shown). Equation (2.15) would appear to be a good model of global mean precipitation for 1905-2004. For all 23 CMIP5 models k_T is positive and significant ($p < 0.05$) and, for all but the

	bcc-csm1-1	CanESM2	CNRM-CM5	CSIRO-Mk3-6-0	MIROC5	MPI-ESM-LR	MRI-CGCM3	NorESM1-M
Equation (2.15)								
k_T (Wm^{-2} K^{-1})	**1.99**	**1.57**	**2.05**	**1.83**	**2.24**	**1.87**	**3.59**	**2.24**
5–95%	**0.11**	**0.16**	**0.44**	**0.29**	**0.24**	**0.40**	**0.58**	**0.20**
β_{LW}	**-0.40**	**-0.47**	**-1.23**	**-0.68**	**-0.53**	**-0.38**	**-0.64**	**-0.59**
5–95%	**0.05**	**0.06**	**0.20**	**0.06**	**0.05**	**0.19**	**0.06**	**0.06**
Equation (2.23)								
k_T (Wm^{-2} K^{-1})	**2.03**	**1.74**	**2.68**	**2.71**	**2.76**	**1.81**	**3.23**	**2.49**
5–95%	**0.16**	**0.35**	**0.73**	**0.19**	**0.39**	**0.88**	**1.01**	**0.33**
β_{LW}	**-0.42**	**-0.55**	**-1.57**	**-1.20**	**-0.74**	-0.35	**-0.48**	**-0.72**
5–95%	**0.08**	**0.17**	**0.38**	**0.09**	**0.14**	0.43	**0.36**	**0.14**
β_{SW}	-0.02	-0.08	-0.27	**-0.46**	**-0.21**	0.02	0.15	-0.11
5–95%	0.06	0.14	0.26	**0.08**	**0.13**	0.26	0.35	0.12

Table 2.5: Regression coefficients for five-year means using just LW forcing data (equation (2.15)) and both LW and SW forcing data (equation (2.23)) for 1905-2004. Values of k_T, β_{LW} and β_{SW} and their 5–95% uncertainty ranges are given. Values is boldface are significant ($p < 0.05$). Results are shown for a subset of 8 of the 23 CMIP5 models listed in Table 2.1.

IPSL-CM5B-LR model, β_{LW} is negative and significant. The mean of β_{LW} across all models is -0.50 $\pm$ 0.33.

To test if there is any value in including net increases in cooling due to changes in tropospheric SW absorbing species concentration as an additional explanatory variable, this model is extended to

$$L\Delta P \simeq k_T \Delta T + \beta_{LW} F_{LW} + \beta_{SW} F_{SW}. \tag{2.23}$$

This extra term, $\beta_{SW} F_{SW}$, was assumed to equal zero previously. Here k_T is positive and significant in all cases, β_{LW} is negative and significant in 21 of 23 CMIP5 models, and β_{SW} is small and generally negative (Table 2.5). It is significant in only 7 of 23 cases. Multi-model mean values of β_{LW} and β_{SW} are -0.62 $\pm$ 0.48 and -0.10 $\pm$ 0.28. Overall, this extra variable is consistent with zero and has little explanatory power. The simpler model, equation (2.15), would appear to be a more suitable model of global mean precipitation.

Repeating the process for the CMIP3 models, using equation (2.15), k_T is positive and significant in 15 of 16 models and β_{LW} is negative and significant in 13 of 16 models. The multi-model mean β_{LW} is -0.44 $\pm$ 0.26. The fact that multi-model mean β_{LW} is similar in CMIP3 and CMIP5 explains why the precipitation offset-aerosol forcing relationship is also similar. Using β_{LW}, as diagnosed from the transient scenario above for CMIP5 models, the precipitation offset can be

estimated with equation (2.22), if F_{SW_2} is assumed to be equal to the strength of aerosol forcing as approximated by the method in Section 2.3.1.1. The predicted precipitation offset is strongly correlated with the actual precipitation offset ($r = 0.90$, $p < 0.01$). This is marginally stronger than the correlation between the precipitation offset and the strength of aerosol forcing alone ($r = 0.80$, $p < 0.01$). The same is true when considering the CMIP3 models.

2.7 Conclusions

It has been shown that a model can be built for global mean precipitation changes. These changes are a function of global mean temperature changes and the climate forcing driving these temperature changes. This is because climate forcings can directly impact the radiative properties of the troposphere independent of changes in temperature, thus changing the tropospheric cooling, which is balanced by changes in tropospheric latent heating associated with changes in global mean precipitation. The temperature dependent term, as in previous studies, is found to be well constrained with a multi-model mean k_T – the sensitivity of tropospheric cooling to ΔT – of 2.25 ± 0.41 Wm^{-2} K^{-1}. This is calculated using idealised experiments. The value of k_T is determined by numerous climate feedbacks within a GCM. Importantly though, SW cloud feedbacks, which cause much of the uncertainty in estimates of ECS, have little effect on the tropospheric energy budget.

The direct effect is far more uncertain and model dependent. Generally, the uncertainty is smaller for GHGs, for which radiative effects – particularly the ratio of TOA forcing to atmosphere forcing – are well understood and radiative forcing time series are considered to be reasonably reliable. The direct effect of sulphate aerosols is considered to be small, or even negligible, because these concern the scattering of SW radiation in the troposphere rather than absorption. Conversely, the direct effect of black carbon aerosols is poorly understood. Some work suggests it can strongly oppose temperature-dependent increases to the extent that the

overall precipitation change is negative. Here, black carbon probably accounts for the poorer performance of the simple precipitation model for some GCMs. Black carbon forcing is difficult to separate from other forcings in the absence of targeted idealised experiments.

A linearised global energy budget approach has been used to calculate the net, LW and SW radiative forcing time series in CMIP5 and CMIP3 models, with the climate feedback parameter (α) calculated using abrupt4×CO_2 experiments. The merits and potential flaws of this approach are discussed in more detail in Sections 2.2 and 5.2.

Sulphate aerosol SW forcing has a larger effect on global mean precipitation than GHG LW forcing per degree of global warming. In a simplified representation of twentieth-century climate change, just these two prominent forcings are considered, with sulphate forcing modelled as a mid-twentieth-century step change and GHG forcing as increasing linearly. This means that the global mean precipitation-temperature relationship can be thought of as two GHG-driven straight lines, offset by sulphate aerosol forcing. The magnitude of this offset is hypothesised to be a function of the increase in mid-twentieth-century aerosol forcing in a GCM (simply referred to as the strength of aerosol forcing) and the reasons for this are explored in Section 2.6.

The global mean precipitation offset, or response to aerosol forcing, is found to be strongly correlated with the strength of aerosol forcing in both CMIP5 and CMIP3. It is expected that the relationship is most robust when considering just models where black carbon is not a significant forcing. This is because TOA black carbon forcing can be positive and surface black carbon forcing negative, while it is expected to increase the magnitude of the precipitation offset if significant. Therefore, for a consistent sign convention – where more negative aerosol forcing corresponds with a more negative precipitation offset – it could be better to diagnose global aerosol forcing at the surface instead. Here, it results in a marginal increase in the offset-forcing correlation. But the advantages of this approach

when studying regional land precipitation in Chapter 3 are more obvious.

The forcing, particularly aerosol forcing, and feedback/sensitivity remain the most uncertain components of the global energy budget. Previous studies have suggested that the range in model simulated twentieth-century temperature changes can be narrowed through correlating the climate feedback parameter and net TOA forcing. It is found that there is still evidence for such a trade off in CMIP5 models. However, with ever more complex representations of, for example, aerosol-cloud interactions (see Section 1.4.2) the radiative forcing in CMIP5 models is even more uncertain, relative to CMIP3. The 5–95% uncertainty range in 2002 SW forcing is larger (by a factor of 2) in CMIP5 than CMIP3. But with little change in the range of diagnosed climate feedback parameters between intercomparison projects, a strong anti-correlation between 2002 SW forcing and 2002 LW forcing is found to produce a similar range in end of twentieth-century net forcing in CMIP5 as in CMIP3.

It appears that CMIP3 and CMIP5 models simulate the observed global mean temperature trend accurately, but for different reasons. However, the precipitation offset would appear to be similar for a given aerosol forcing in the two model ensembles despite obvious differences in SW and LW forcing. The simple representation of twentieth-century climate is revisited and the precipitation offset is found to be a function of the increase in mid-twentieth-century aerosol forcing and a coefficient that determines the ratio of atmosphere LW to TOA LW forcing. The multi-model mean and range of this coefficient is found to be similar in CMIP3 and CMIP5.

Chapter 3

Northern Hemisphere land precipitation

In the last chapter it was shown how changes in global mean precipitation are constrained by the tropospheric energy budget and how a marked change in mid-twentieth-century aerosol forcing can lead to a significant decrease in precipitation – a decrease greater than would be expected were global mean precipitation simply a function of global mean temperature. While some of the methods used in Chapter 2 relied on simple representations of twentieth-century climate change, there is nevertheless compelling evidence of a global mean precipitation response to sulphate aerosol forcing. Frustratingly, this cannot be verified by global mean precipitation observations for reasons outlined in Section 1.2. But suitable observations do exist over the global land area, particularly the Northern Hemisphere land area. However, characterising regional precipitation change is beset with difficulties (see Section 1.3.2). Further, past regional forcing is even more uncertain. It might still be the case, though, that some robust precipitation change features are evident.

3.1 Patterns of regional climate change

The NHML land region has been the source for a large proportion of global emissions originating from human activity (Lamarque et al., 2010), including short-lived forcing agents such as aerosols. Our longest, most comprehensive temperature and precipitation observations also exist here (see Section 1.2). These have allowed for the identification of a temperature response to aerosol forcing in observations and climate models (Stott, 2003b).

Throughout this chapter, all precipitation and temperature model data are masked to observed datasets. The four gridded observational precipitation datasets outlined in Section 1.2.1.4 – GHCN, Zhang, CRU and GPCC – are all considered, although the GHCN dataset is used in all main analyses, unless stated otherwise. The CRU and GPCC datasets are spatially interpolated, but here only grid boxes where real observations exist are considered. These data (CRU and GPCC) are gridded to the same 5° × 5° grid to be consistent with the GHCN and Zhang datasets, as well as the Met Office Hadley Centre-Climate Research Unit version 4 (HadCRUT4) temperature observations (Morice et al., 2012). The four precipitation datasets are not masked to be spatially and temporally consistent with each other. Different datasets contain different station records, with some favouring fewer long-term homogenised records and others selecting a much greater number of short-term records, thus testing the sensitivity of the analyses to varying spatial and temporal coverage (Polson et al., 2013, see Section 1.2.1.4). For the HadCRUT4 gridded global surface temperature anomalies dataset the median of the 100 ensemble members available is used.

All model data are first regridded to the common 5° × 5° grid. For simulated precipitation, each of the four observed datasets is masked in turn. The only exception to this is when global mean precipitation is considered, such as in Figures 3.1a and 3.6. Here, unmasked global means are shown instead. Simulated temperature is masked to the HadCRUT4 dataset (land and ocean) in all instances.

Anomalies are with respect to the period 1961-1990 (the years with most available observations). Again, Figures 3.1 and 3.6 are exceptions with anomalies instead with respect to the mean of a pre-industrial control simulation.

3.1.1 The precipitation-temperature relationship regionally

Figure 3.1a shows the five-year global mean precipitation-temperature relationship for three twentieth-century experiments driven with different forcings using the CanESM2 climate model. Both the experiment forced only by GHGs and the anthropogenic aerosol alone experiment have a linear precipitation-temperature relationship, but the change in precipitation per unit change in temperature is greater in the latter. The all forcings experiment reflects the temporal evolution of twentieth-century GHG and anthropogenic aerosol forcing, with two GHG-driven straight lines offset by mid-twentieth-century aerosol-driven cooling (see Section 2.3).

Mid-twentieth-century aerosol emissions induced an interhemispheric forcing asymmetry that has been accredited with the southward shift of the tropical rain belt seen in models and observations (Chang et al., 2011; Hwang et al., 2013), drying the NHT region and wetting the Southern Hemisphere tropics. Therefore, the five-year NHT land mean precipitation-global mean temperature relationship (Figure 3.1b) is comparable to that in Figure 3.1a in all three experiments, demonstrating a precipitation response to largely remote NHML aerosol forcing. However, aerosol forcing is also expected to decrease the local hydrological cycle in most situations, by cooling the surface and encouraging local atmospheric subsidence (sulphate) or through increasing atmospheric radiative absorption (black carbon). Indeed, idealised experiments have demonstrated that local aerosol forcing can cause large circulation changes and an associated reduction in NHML precipitation (Shindell et al., 2012). Modelled NHML land mean precipitation shares a similar relationship with global mean temperature as both global mean

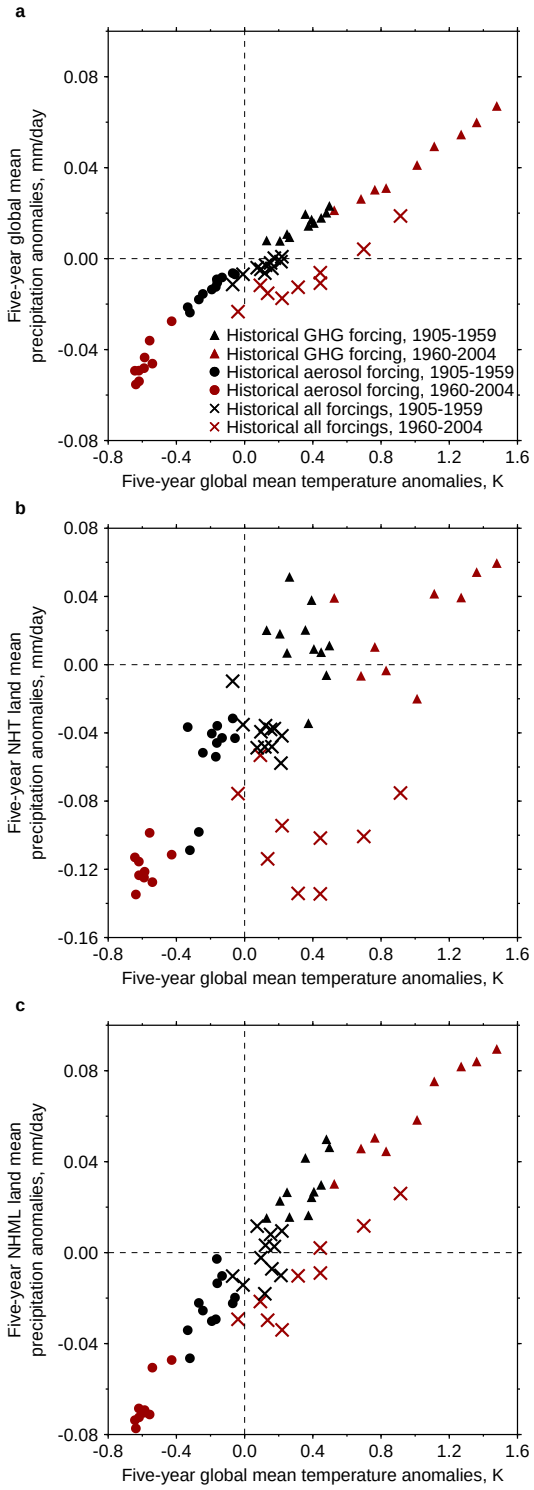

Figure 3.1: Five-year global mean precipitation-temperature relationships (a), five-year NHT land mean precipitation-global mean temperature relationships (b) and five-year NHML land mean precipitation-global mean temperature relationships (c) with the CanESM2 climate model. Three twentieth-century experiments are included – one driven by GHG forcings alone, one driven by anthropogenic aerosol forcings alone and one driven by all forcings. Five members contribute towards the ensemble mean of each experiment, with anomalies given relative to the mean of a pre-industrial control simulation. Temperature is masked to HadCRUT4 observations (see Section 3.1). Precipitation in (b) and (c) is masked to GHCN observations (see Section 3.1).

precipitation and NHT land mean precipitation (Figure 3.1c), so evidence for an aerosol response in NHML land precipitation observations is expected.

3.1.2 Identifying potential aerosol signatures

Figure 3.2a shows the surface temperature gradient anomalies from the HadCRUT4 dataset (Morice et al., 2012) and output from historically forced runs of 23 climate models (Table 2.1). Here, mean temperature gradient (with land and ocean grid boxes considered) is defined as the difference between NHML surface temperature anomalies and Southern Hemisphere extratropical (SHEXT) surface temperature anomalies. The NHML region is defined as the latitude band bounded by 30°N and 65°N. However, the findings are largely insensitive to slight shifts in these bounds. The SHEXT region is the latitude band bounded by 30°S and 90°S. Temperature anomalies in the SHEXT region show continual warming in the absence of significant local aerosol forcing; thus, the gradient change accentuates the temperature response to aerosols in the NHML region. The observations and most models show a mid-twentieth-century hiatus or downward trend, but the models do not simulate the observed abrupt decrease in the late 1960s (Friedman et al., 2013), which is widely attributed to internal variability (Thompson et al., 2010). Significant warming again dominates in the late twentieth century.

Most models and GHCN dataset precipitation observations show a NHT land mean mid-twentieth-century drying, followed by a late-twentieth-century wetting (Figure 3.2b). This telltale aerosol signature is also evident when considering all Northern Hemisphere land masses (Wild, 2012; Wu et al., 2013), with the NHT land region driving the trends (Figure 3.3). There is no overall twentieth-century trend in model simulated NHML land mean precipitation anomalies (Figure 3.2c), consistent with simulated NHT land mean precipitation anomalies and simulated global mean precipitation anomalies (Figure 3.4). However, there is a positive trend in observed NHML land mean precipitation anomalies, with no evidence of a mid-twentieth-century drying signal. Three other gridded observational pre-

cipitation datasets (see Section 1.2.1.4) are used to check for consistency. All four datasets show a near-identical positive trend (Figure 3.5) and no evidence for an aerosol-driven drying. No uncertainty estimates are shown in Figure 3.5, because none are provided with the respective datasets. However, much like the HadCRUT4 dataset (Morice et al., 2012), confidence intervals are expected to be wider at the start of the twentieth century when spatial coverage of the rain gauge network was poor.

3.2 The expected response to aerosol forcing

Changes in mean temperature gradient, NHT land mean precipitation and NHML land mean precipitation are seemingly more sensitive to spatially heterogeneous (mostly NHML) aerosol forcing than more homogeneous GHG forcing. But to what extent does the varying strength of NHML aerosol forcing, in models and the real world, manifest itself in the magnitude of the departure from an otherwise linear GHG-driven response? Is the offset approach suitable for the mean temperature gradient and regional land mean precipitation?

3.2.1 Temperature gradient

The work of Friedman et al. (2013) suggested that the interhemispheric temperature asymmetry – the temperature contrast between the Northern and Southern Hemispheres – showed no significant trend over most of the twentieth century, before an obvious upward trend around 1980. This was attributed to the significant warming of the Arctic and Northern Hemisphere landmasses as GHG emissions accelerated and sulphate aerosol emissions decreased due to clean air policies. Here, a subtly different pattern is found if defining the temperature gradient as NHML temperature minus SHEXT temperature. The NHML warms slightly more than the SHEXT at the start of the twentieth century, less – or perhaps even cooling – after the Second World War (even if removing the abrupt decrease in the late

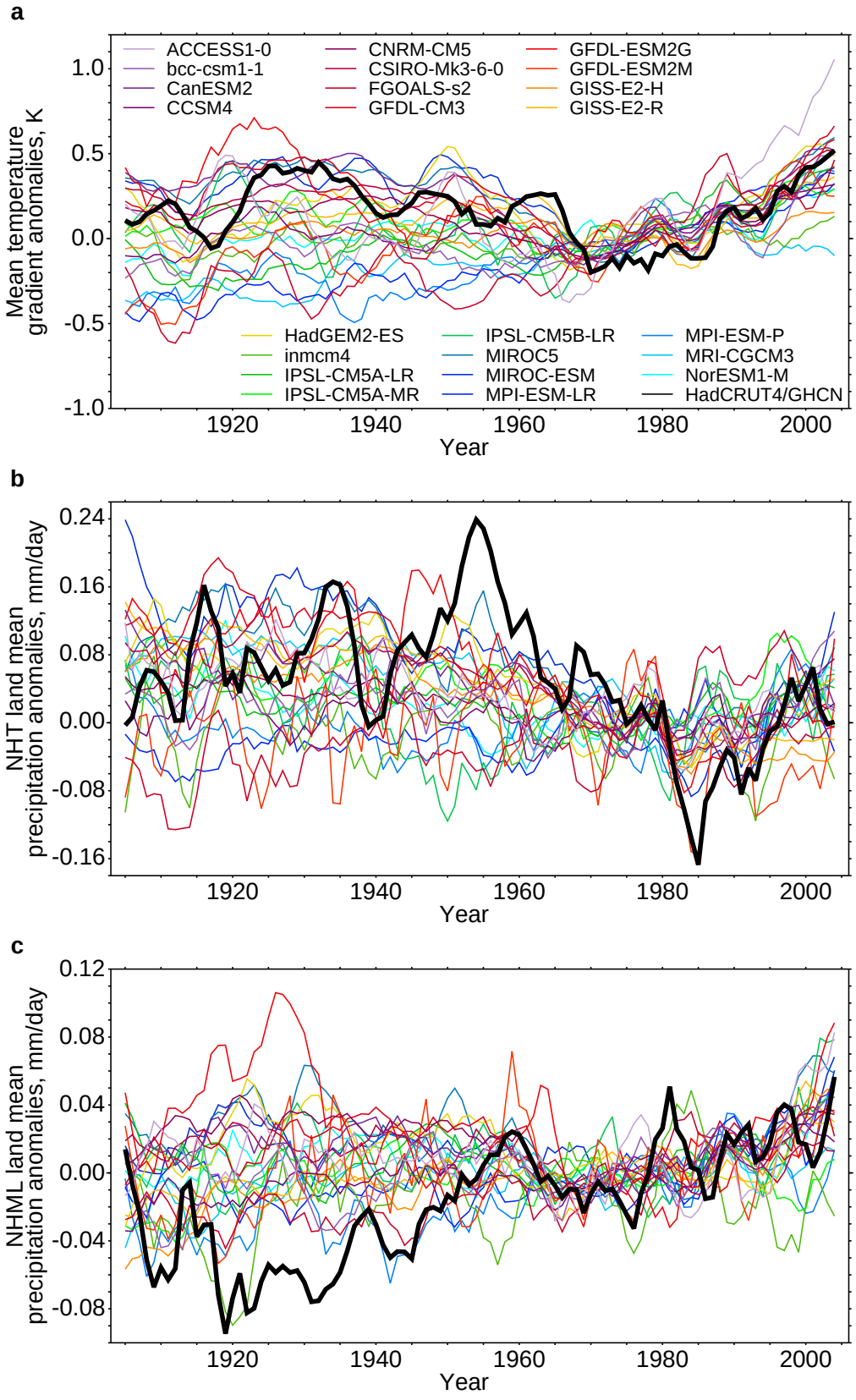

Figure 3.2: Five-year running mean temperature gradient anomalies (a), five-year running mean NHT land mean precipitation anomalies (b) and five-year running mean NHML land mean precipitation anomalies (c). Model data (coloured lines) are from 23 CMIP5 models (Table 1 2.1) and observed data (black lines) are from the HadCRUT4 and GHCN datasets for temperature and precipitation respectively. All model data are masked to the observations and anomalies are with respect to the period 1961-1990.

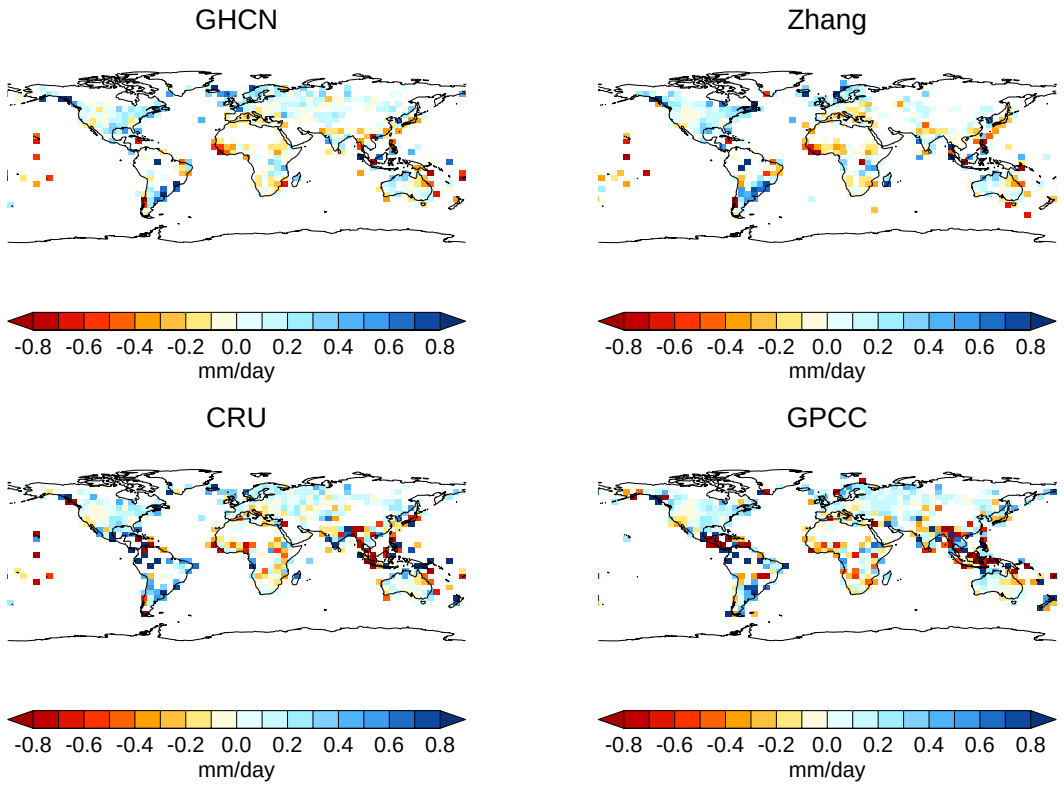

Figure 3.3: Land precipitation change between 1905-1924 and 1985-2004 (mm day^{-1}) for the GHCN, Zhang, CRU and GPCC datasets (see Section 1.2.1.4). Only grid boxes where data exists for at least 50% of months in each twenty-year period are considered.

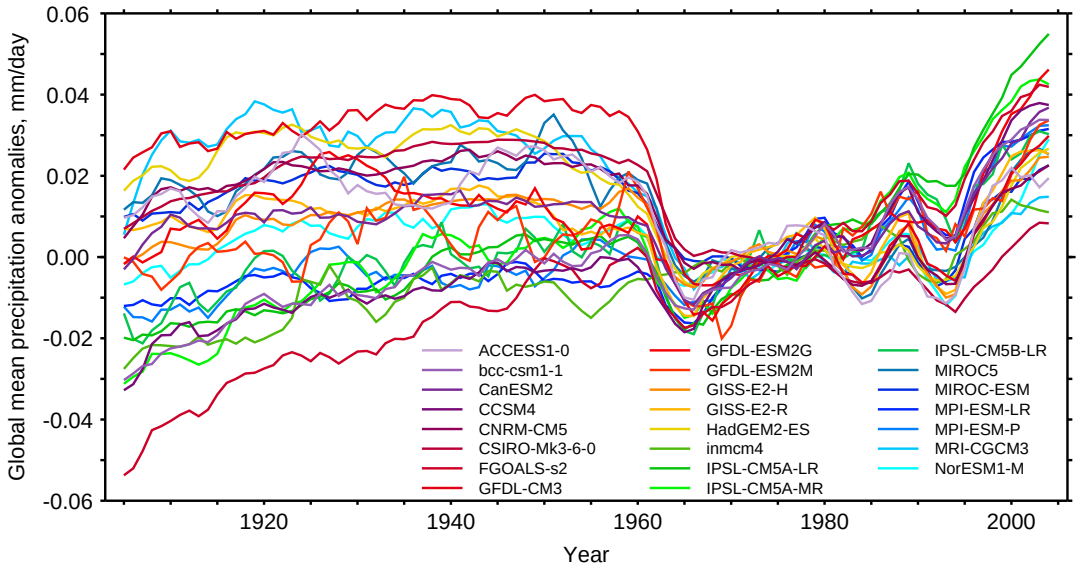

Figure 3.4: Five-year running mean global mean precipitation anomalies. Data are from 23 CMIP5 models (Table 2.1). All data are for 1905-2004 and anomalies are with respect to the period 1961-1990.

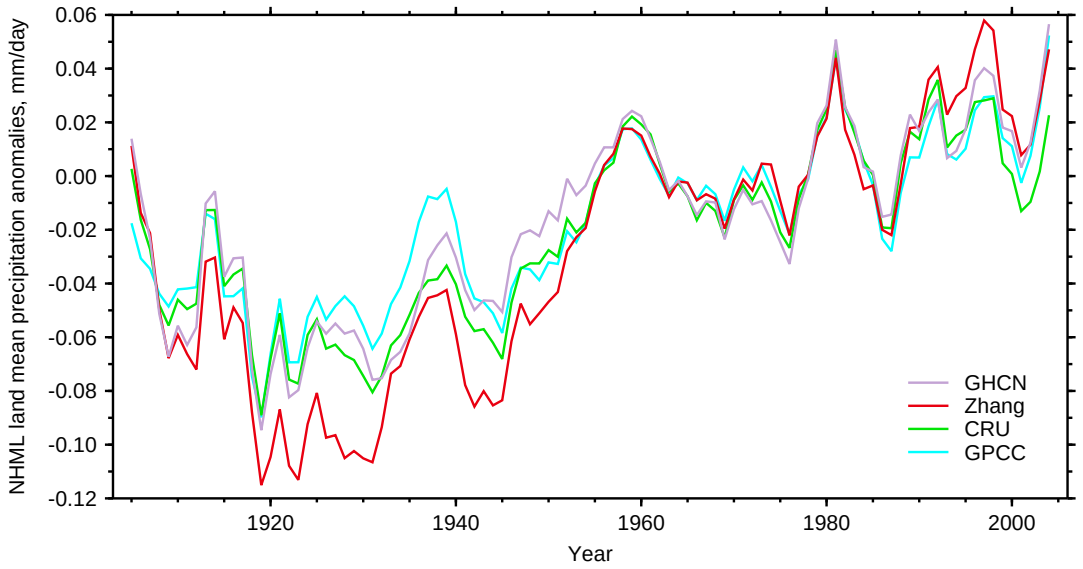

Figure 3.5: Five-year running mean NHML land mean precipitation anomalies. Data are from four gridded observational datasets (see Section 1.2.1.4). All data are for 1905-2004 and anomalies are with respect to the period 1961-1990.

1960s – see Section 3.3.2) and warms noticeably more rapidly after about 1975 (Figure 3.2a). Shindell and Faluvegi (2009) found the same differences in warming and cooling when using a near-identical definition for the meridional temperature gradient. The authors found that GHGs, ozone and internal variability alone could not account for the observed temperature trends. But through imposing radiative forcing from sulphate aerosols in space and time, which roughly agreed with the best historical emissions estimates available, these differences were explained.

Through defining the temperature gradient as the NHML-SHEXT difference, the five-year mean temperature gradient-global mean temperature relationship closely resembles the historical all forcings patterns in Figure 3.1 (not shown). Typically, in observations and models, the post-1960 data are strongly correlated as the GHG-driven and global brightening signal emerges. But a linear fit to these data is clearly offset from the pre-1960 data due to the marked increase in mid-twentieth-century aerosol forcing. Therefore, it is expected that the offset measure in Section 2.3.1.2 is suitable for measuring the magnitude of the gradient temperature response to aerosol forcing. Further, the size of the offset should be dependent on the strength of aerosol forcing.

3.2.2 NHT land precipitation

Similarly, the NHT land mean precipitation response to remote NHML aerosol forcing should be detectable as an offset. It is likely that the overall NHT land precipitation decrease to aerosol forcing will be made up of two components. The first should broadly scale with global mean precipitation changes (Figure 3.6). As with global mean precipitation changes, radiative forcings that cause a given change in tropical/global temperature can produce different changes in mean tropical precipitation (Ferraro et al., 2014). This is because the tropical circulation is weakened most by climate forcings that cause the most tropospheric heating, which increases tropospheric static stability and suppresses convection.

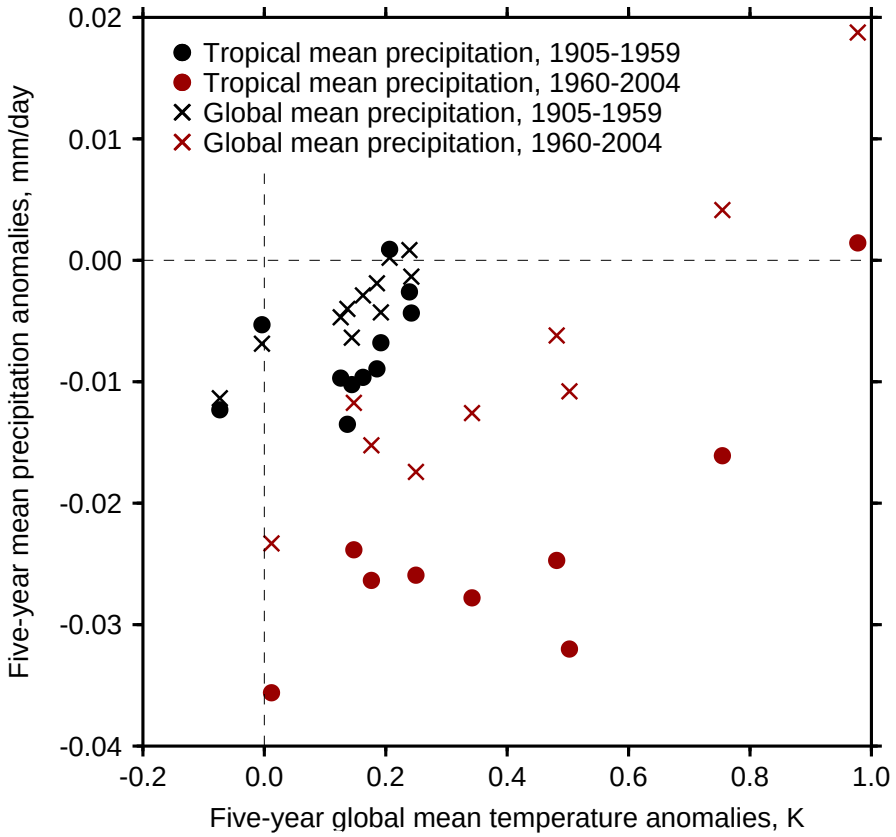

Figure 3.6: Five-year global mean precipitation-temperature relationship (crosses) and five-year tropical (land and ocean) mean precipitation-global mean temperature relationship (filled circles)) in the historical all forcings experiment with CanESM2. Five members contribute towards the ensemble mean, with anomalies given relative to the mean of a pre-industrial control simulation. No masking is applied (all grid boxes are considered).

As discussed in Section 1.4.4, the temperature gradient induced by mid-twentieth-

century NHML aerosol forcing was probably the cause of a southward shift in the tropical rain belt seen in models and observations around this time. This should further decrease precipitation at this time and enhance the precipitation offset. However, because aerosol emissions were significantly reduced in the 1980s and 1990s in North America and Europe, tropical precipitation then migrated northwards, bringing to an end the severe Sahel drought of the 1970s (Hwang et al., 2013). This would act to increase NHT land mean precipitation. It is clear that the NHT land precipitation response is somewhat complex. Explosive volcanic eruptions can also act to shift the tropical rain belts and there is even evidence for this in river discharge records in tropical Africa (Oman et al., 2006). The ability to detect a robust precipitation response is further hampered by the sparse in situ precipitation network over tropical land, to which all model data are masked (see Section 3.1). Therefore, NHT land mean precipitation changes are likely to be noisy, but it is predicted that a robust aerosol response should still be detectable.

3.2.3 NHML land precipitation

It is found that NHML land mean precipitation changes are remarkably similar to global mean precipitation changes in a given CMIP5 model (Figure 3.1), similar to the findings of Wu et al. (2013). This suggests that the contrast in GHG and sulphate aerosol direct effects can cause a significant aerosol response in local NHML land precipitation. Therefore, the offset for each of the three variables discussed above is calculated using the method outlined in Section 2.3.1.2. This section also serves as a reference point for information regarding uncertainty calculations.

3.2.4 Diagnosing NHML surface aerosol forcing

In Chapter 2 the merits of using surface aerosol forcing rather than TOA aerosol forcing were discussed. Diagnosing aerosol forcing at the TOA or surface made

little difference when considering global means, but is expected to be more important regionally, especially when considering the local precipitation response to NHML aerosol forcing. Significant black carbon forcing is likely to stabilise the tropospheric profile and suppress convection, through SW tropospheric heating (positive forcing at the TOA and negative forcing at the surface). Therefore, surface aerosol forcing is used here.

Forcing time series are calculated using the linear forcing-feedback model outlined in Section 2.2. But changes in mean temperature gradient, NHT land mean precipitation and NHML land mean precipitation are expected to respond most obviously to mid-twentieth-century NHML aerosol forcing, rather than global aerosol forcing. Therefore, NHML forcing is calculated by regressing the regional average surface energy imbalance against the global mean surface air temperature. Section 2.3.1.1 studied the suitability of using an indirect approach for calculating the strength of aerosol forcing (through using SW forcing as a proxy) in a temporal sense. Here, Figure 3.7 evaluates the suitability spatially, albeit for only one CMIP5 model (CanESM2). Only anthropogenic aerosol forcings are time-varying over the historical period in the top two panels. The similarity of the strength of surface forcing in these two panels suggests that forcing at SW frequencies accounts for the vast majority of net aerosol forcing (at least in this model).

The bottom panel, which includes a representation of all major climate forcings, shows many of the same features, especially over the NHML region. These include the strong surface aerosol forcing signature over parts of southeast Asia and China and weak, or even positive, surface aerosol forcing over areas of northeast North America and northwest Europe, perhaps highlighting where clean air policies have had the greatest impact. But other forcings (and feedbacks) will undoubtedly impact the geographic distribution of surface SW forcing in the bottom panel. For example, ozone net radiative forcing is significant, especially tropospheric ozone forcing (Figure 1.3). But the LW component still dominates global mean pre-industrial to present day net tropospheric ozone forcing and is roughly four

times greater than the SW component according to IPCC AR5 estimates (Myhre et al., 2013).

However, some differences are evident in the panels representing surface SW forcing for the contrasting experiments (top right and bottom), especially in the Southern Hemisphere. The Southern Hemisphere mean surface SW forcing in the all forcings experiment is -0.53 Wm^{-2} compared with -0.76 Wm^{-2} in the anthropogenic aerosols only experiment, perhaps due to a positive contribution from increasing ozone concentrations in the former. This is of course complicated by SW radiative forcing due to stratospheric ozone depletion over the Southern Hemisphere polar region, which is believed to make a positive contribution at high latitudes (Myhre et al., 2013). But because the strength of surface SW forcing is calculated as the difference between 1960-2004 mean forcing and 1905-1959 mean forcing, forcing due to changes in stratospheric ozone will probably be minimal in the bottom panel of Figure 3.7. This is because the marked depletion and subsequent recovery of Southern Hemisphere polar stratospheric ozone occurred after 1960.

The impact of water vapour may also be evident in Figure 3.7. The specific humidity is mostly dependent on air temperature (see Section 1.3.1) rather than emissions, so can be considered a feedback rather than a forcing. It does however have a strong greenhouse effect. Therefore, as surface air temperature increases in the all forcings experiment (bottom panel) water vapour could impact surface SW fluxes. Decreases in clear-sky downward surface SW flux associated with water vapour increases and atmospheric SW absorption under a scenario of increasing CO_2 concentrations might be expected (Takahashi, 2009). This might have some role in explaining the negative surface SW forcing in the Northern Hemisphere polar region in the all forcings experiment (bottom panel), which is less evident in the anthropogenic aerosol forcings only experiment (top panels). This might seem counterintuitive since significant sea ice loss (a positive feedback) post-1960 should lead to a lower surface albedo in this region and a positive downward

surface SW flux. But this region has seen some of the largest relative increases in near-surface specific humidity, due to significant near-surface warming associated with Arctic amplification (Screen and Simmonds, 2010) and additional moisture because of enhanced evaporation due to sea ice loss (Screen et al., 2013).

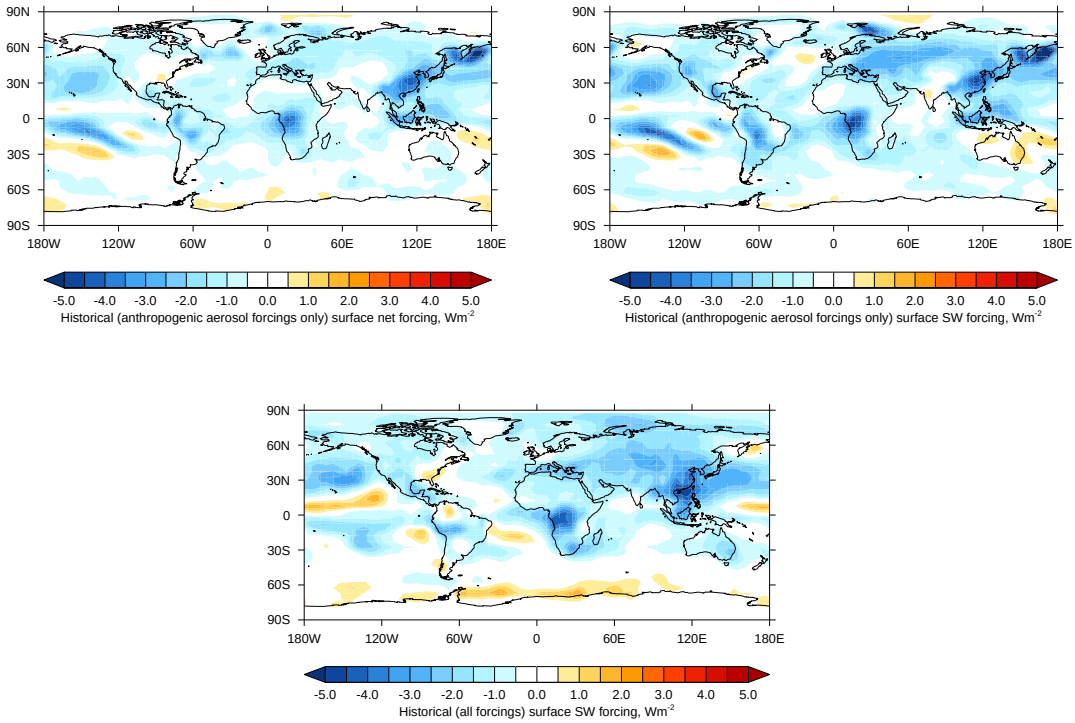

Figure 3.7: The strength of surface net forcing (top left) and surface SW forcing (top right) in the historical (anthropogenic aerosol forcings only) experiment and surface SW forcing (bottom) in the historical (all forcings) experiment for the CanESM2 model. Local (net or SW) forcing time series are diagnosed by regressing the local surface energy imbalance (due to net or SW radiative forcing) against the global mean surface air temperature with the strength of (net or SW) forcing then calculated using the methodology outlined in Section 2.3.1.1.

The spatial pattern of the standard deviation of surface SW forcing (as calculated in Section 2.3.1.1) in CMIP5 is shown in Figure 3.8. The greatest model spread is seen in regions of strong surface SW forcing (compare the bottom panel of Figure 3.7 with Figure 3.8). This regional decomposition was suggested by Boer and Yu (2003), although the results here are insensitive to regressing against regional surface air temperature or global surface air temperature. The components of the surface energy budget also include the turbulent heat fluxes, latent heat and sensible heat, although here only SW surface radiative fluxes are considered.

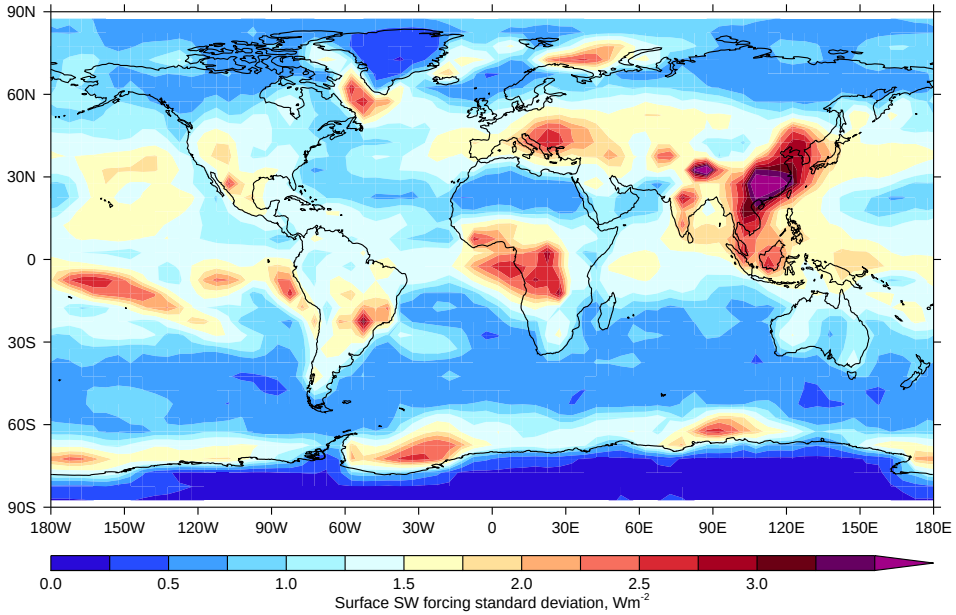

Figure 3.8: The standard deviation of the strength of surface SW/aerosol forcing in CMIP5. Local SW forcing time series are diagnosed by regressing the local surface energy imbalance (due to the SW component) against the global mean surface air temperature with the strength of SW/aerosol forcing then calculated using the methodology outlined in Section 2.3.1.1. For each grid box the standard deviation of the surface SW forcing is then taken.

As in Section 2.3.1.1, SW surface forcing from a historical all forcings experiment is taken to represent surface aerosol forcing. NHML surface SW forcing from a historical all forcings experiment is largely representative of NHML surface net radiative forcing from a historical anthropogenic aerosol forcings alone experiment (not shown), particularly when considering the difference between two long-term means. Model simulated radiative fluxes are regridded to the common $5° \times 5°$ grid. The strength of NHML surface aerosol forcing in models is diagnosed by taking the difference between average NHML mean (land and ocean) surface SW forcing before and after 1960, as described in Section 2.3.1.1, where details about uncertainty range calculations can also be found.

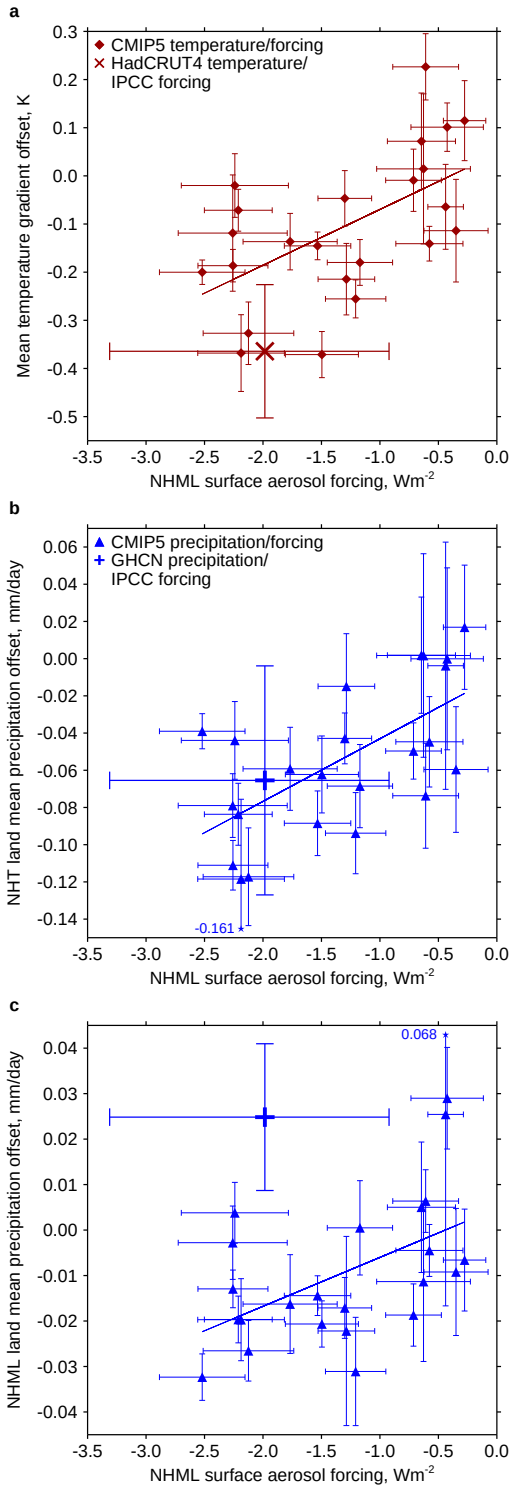

Figure 3.9: Mean temperature gradient offset against NHML surface aerosol forcing (a), NHT land mean precipitation offset against NHML surface aerosol forcing (b) and NHML land mean precipitation offset against NHML surface aerosol forcing (c). Error bars represent the 5–95% uncertainty range (see Sections 2.3.1.1 and 2.3.1.2), with IPCC equivalent NHML surface aerosol forcing and the associated 5–95% uncertainty range used for the observed data points (see Section 3.3.1 and Figure 3.11). The values next to the error bars in b and c indicate where the uncertainties extend beyond the scale.

3.3 The aerosol response in models and observations

The temperature gradient (Figure 3.9a), NHT land precipitation (Figure 3.9b) and NHML land precipitation (Figure 3.9c) offsets are all correlated with NHML surface aerosol forcing ($r = 0.57$, $r = 0.64$ and $r = 0.51$ respectively, $p < 0.05$). The land precipitation offset-aerosol forcing correlation is largely independent of the observed precipitation dataset chosen, which provide temporally varying spatial masks for the model simulated output (Figure 3.10). Using NHML TOA aerosol forcing, instead of NHML surface aerosol forcing, has little little impact on the temperature gradient and NHT land precipitation correlations, but the NHML land precipitation correlation does drop to $r = 0.39$, which is not significant at the 0.05 level.

3.3.1 An IPCC equivalent NHML surface aerosol forcing

In IPCC AR5 the present day global mean TOA total aerosol effect is estimated. But an equivalent IPCC value for NHML surface total aerosol effect is required for this chapter to determine whether the observed temperature and precipitation offsets are in keeping with aerosol forcing in this region. Because aerosol over the NHML region contributes a large fraction of the global mean aerosol forcing (Shindell, 2014), the latter should be strongly dependent on the former in GCMs. Therefore, in the presence of a robust relationship, the real world, IPCC equivalent, NHML surface aerosol forcing most likely to have generated the IPCC AR5 present day global mean TOA aerosol forcing can be determined.

The latest IPCC present day (2011 relative to 1750) global mean aerosol forcing estimate is -0.9 [-1.9 to -0.1] Wm^{-2} (Boucher et al., 2013) (see Section 1.4.3). Here, the global mean SW TOA forcing time series, shown in Figure 2.3 (bottom panel), are used to determine an analogous present day (note; 2002 (2000-2004

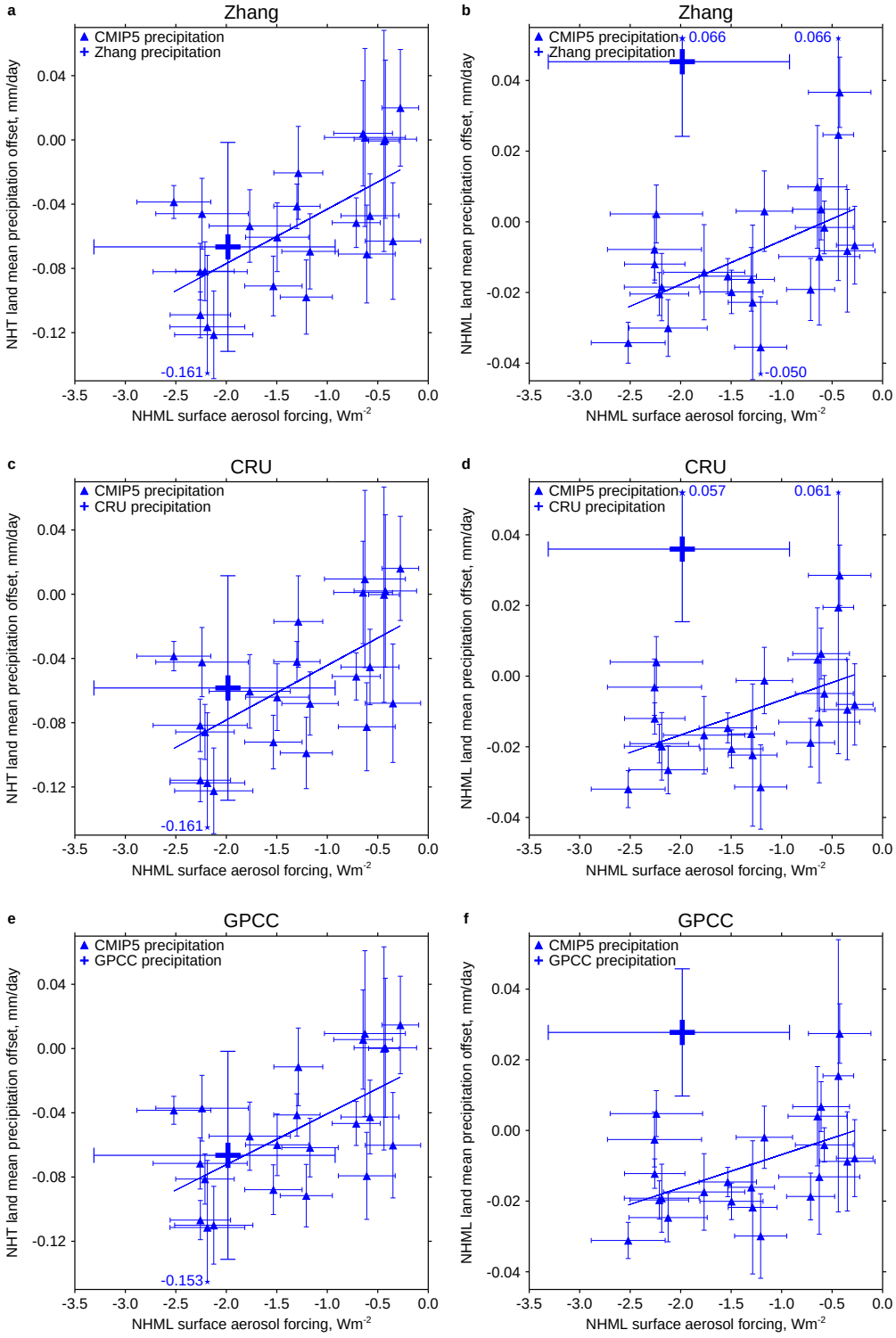

Figure 3.10: NHT land mean (left column) and NHML land mean (right column) precipitation offsets against NHML surface aerosol forcing using masks from the Zhang (top row), CRU (middle row) and GPCC (bottom row) datasets. Error bars represent the 5–95% uncertainty range (see Sections 2.3.1.1 and 2.3.1.2), with IPCC equivalent NHML surface aerosol forcing and the associated 5–95% uncertainty range used for the observed data points (see Section 3.3.1 and Figure 3.11). The values next to the error bars are where uncertainties extend beyond the scale.

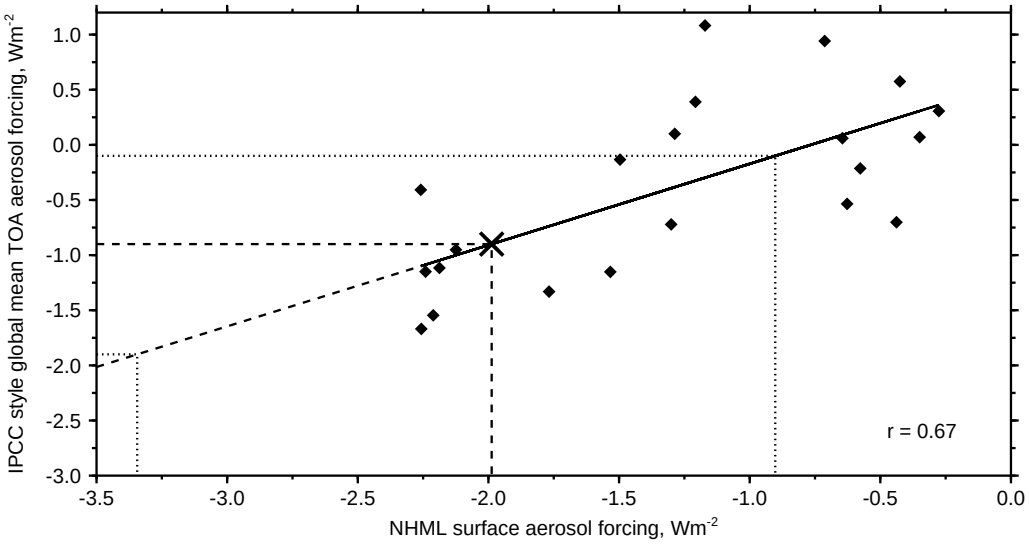

Figure 3.11: IPCC style global mean TOA aerosol forcing against NHML surface aerosol forcing for 21 CMIP5 models (minus the CNRM-CM5 and MRI-CGCM3 models – see text). Using linear regression the IPCC equivalent NHML surface aerosol forcing (cross and dashed lines) and 5–95% uncertainty range (dotted lines) are shown, given the IPCC AR5 present-day global mean aerosol forcing estimate and uncertainty range (-0.9 [-1.9 to -0.1] Wm^{-2}).

average) relative to a pre-industrial control run) global mean TOA aerosol forcing estimate for individual CMIP5 models, referred to as an "IPCC style" estimate in the y-axis of Figure 3.11. This method has some flaws, namely using SW forcing as a proxy for total aerosol forcing, but should be reasonably representative of the relative strength of present day aerosol forcing in the CMIP5 models.

Global mean TOA aerosol forcing estimate is strongly correlated ($r = 0.67$, $p < 0.01$) with NHML surface aerosol forcing, excluding the MRI-CGCM3 and CNRM-CM5 models (Figure 3.11). Considering the spatial pattern of SW forcing in MRI-CGCM3 (not shown), strong aerosol forcing is evident in the tropics, but weak aerosol forcing in the NHML region, contrary to output from other models and our scientific understanding of twentieth-century aerosol forcing. This was also found by Cherian et al. (2014). This model is retained in the main analyses considering the offset-aerosol forcing relationship, because both the mean temperature gradient response and NHML land mean precipitation response are in keeping with weak NHML aerosol forcing. This unusual spatial pattern of SW forcing might be expected to affect the NHT land mean precipitation response,

but no evidence is found for this, consistent with other work, which suggests that tropical precipitation is more sensitive to remote aerosol forcing than local aerosol forcing (Shindell et al., 2012). Further, this would suggest that NHML land mean precipitation appears insensitive to remote tropical aerosol forcing, at least in the MRI-CGCM3 model.

Despite very strong NHML surface aerosol forcing, the CNRM-CM5 model has weak NHML TOA aerosol forcing, so is excluded here (see Section 2.4.1). This is likely to be because of significant black carbon aerosol forcing. Again, this model is included in the main analyses, because significant black carbon forcing would be expected to produce strongly negative NHML surface aerosol forcing and an even greater negative precipitation offset in the mid twentieth century. IPCC equivalent NHML surface aerosol forcing is approximated as -2.0 [-3.3 to -0.9] Wm^{-2} (Figure 3.11).

It should be remembered that the strength of NHML surface aerosol forcing is calculated as the difference between mean NHML mean surface SW forcing before and after 1960. Therefore, models that simulate stronger forcing efficiencies and stronger radiative effects will have a more negative NHML surface aerosol forcing. Cherian et al. (2014) found that these models also simulate stronger solar brightening trends over Europe since the 1990s. Models with greater solar brightening trends were found to have more negative global mean TOA aerosol forcing. Therefore, models with a stronger global dimming trend and, in turn, a more negative NHML surface aerosol forcing, also have a stronger global brightening trend. But the magnitude of the dimming trend – the increase in aerosol forcing around 1960 – would appear to be more important in determining present day global mean TOA aerosol forcing.

3.3.2 Consistency between models and observations

Using the temperature gradient offset-aerosol forcing regression coefficients, the observed temperature offset of -0.36 ± 0.14 K corresponds to an aerosol forcing of -3.5 ± 1.2 Wm^{-2} (Figure 3.9a), compared with the IPCC equivalent NHML surface aerosol forcing of -2.0 [-3.3 to -0.9] Wm^{-2}. The difference is a consequence of the likely ocean circulation-driven abrupt decrease in the observed mean temperature gradient in the late 1960s (Thompson et al., 2010). If this abrupt decrease is not captured by CMIP5 models (e.g. Friedman et al. (2013)) the agreement should be better if it is removed from observations.

The difference between the 1963-67 and 1973-77 mean temperature gradient means (0.36 K) is taken as the magnitude of the late 1960s shift. Increasing all 1968-2004 annual anomalies by this value, a new offset for the mean temperature gradient with this abrupt shift removed is calculated. A corrected observed temperature offset more indicative of a response to just anthropogenic aerosol forcing of -0.15 ± 0.04 K gives an aerosol forcing of -1.7 ± 0.4 Wm^{-2}. The observed NHT land precipitation offset agrees well with the IPCC equivalent NHML surface aerosol forcing. Here, the offset of -0.065 ± 0.062 mm day^{-1} gives an aerosol forcing of -1.7 ± Wm^{-2} (Figure 3.9b).

The simple technique to remove the abrupt temperature shift gives a mean temperature gradient offset that is in better agreement with what might be expected from real-world NHML surface aerosol forcing. However, it is also possible that part of the observed NHT land mean precipitation offset is attributable to this sudden shift in the mean temperature gradient. While this might influence the robust relationship between the observed NHT land mean precipitation offset and IPCC equivalent NHML surface aerosol forcing, it is again noted that a strong link between mid-twentieth-century aerosol forcing (primarily remote forcing from the NHML region) and a southward tropical precipitation shift (comparable to decreasing NHT land mean precipitation) has been found in other studies (Chang et al., 2011; Hwang et al., 2013). While a robust feature, the southward precip-

	Expected	Actual
Temperature gradient (K)	-0.18 [-0.34 to -0.06]	-0.36 ± 0.14
Temperature gradient no abrupt shift (K)	-0.18 [-0.34 to -0.06]	-0.15 ± 0.04
NHT land precipitation (mm day^{-1})	-0.076 [-0.122 to -0.040]	-0.065 ± 0.062
NHML land precipitation (mm day^{-1})	-0.017 [-0.031 to -0.005]	+0.025 ± 0.016

Table 3.1: Expected and actual offsets for mean temperature gradient (with and without the abrupt late 1960s shift), NHT land mean precipitation and NHML land mean precipitation. The expected offset is calculated given the offset-aerosol forcing regression coefficients (from the fit to the CMIP5 models) and the IPCC equivalent NHML surface aerosol forcing (see Section 3.3.1). The actual offset is that calculated from observed data.

itation shift – defined using a different measure – has actually been found to be underestimated in CMIP5 models (Hwang et al., 2013). Further, it has been suggested that aerosol emissions may be the dominant influence on North Atlantic ocean variability (Booth et al., 2012), which is frequently identified as the key contributor to the abrupt mean temperature gradient shift, casting some doubt on whether this is simply an internal variability-driven signal.

The observed NHML land precipitation offset of $+0.025 \pm 0.016$ mm day^{-1} seems to contradict both the observed temperature gradient offset and the observed NHT land precipitation offset, while equivalent to an aerosol forcing of $+1.5 \pm 1.6$ Wm^{-2}, calculated using the NHML land precipitation offset-aerosol forcing regression coefficients (Figure 3.9c). This is clearly inconsistent with the IPCC equivalent NHML surface aerosol forcing (also see Table 3.1). This could imply that the correlation across CMIP5 models is not robust. It is noted, however, that the observed NHML land precipitation offset is likely to be more positive than all but two models masked to the GHCN dataset when considering the four observed datasets (Figures 3.9c and 3.10).

A stronger correlation is found with the removal of the GISS-E2-H and GISS-E2-R climate models ($r = 0.65$, $p < 0.01$). Surface LW and surface aerosol forcing are clearly anti-correlated in the NHML region ($r = -0.72$, $p < 0.01$), but the GISS-E2-H and GISS-E2-R climate models have anomalously strong NHML surface LW forcing (Figure 3.12). Because of this large increase in NHML surface LW forcing in the mid twentieth century, these models do not show the interhemispheric forc-

ing asymmetry expected from their level of NHML surface aerosol forcing. In turn, both the mean temperature gradient and the NHT land mean precipitation offsets will not be as great as implied by NHML surface aerosol forcing alone. Likewise, the NHML land mean precipitation offsets are underestimated, as SW-induced surface cooling is counteracted by significant LW-induced surface warming. Shindell (2014) also excluded the GISS-E2-R model from a study looking into the impact of hemispherically asymmetric forcing. The author found that its Northern Hemisphere and Southern Hemisphere forcings were quite similar, but instead attributed this to the model overestimating negative Southern Hemisphere nitrate aerosol and ozone forcing.

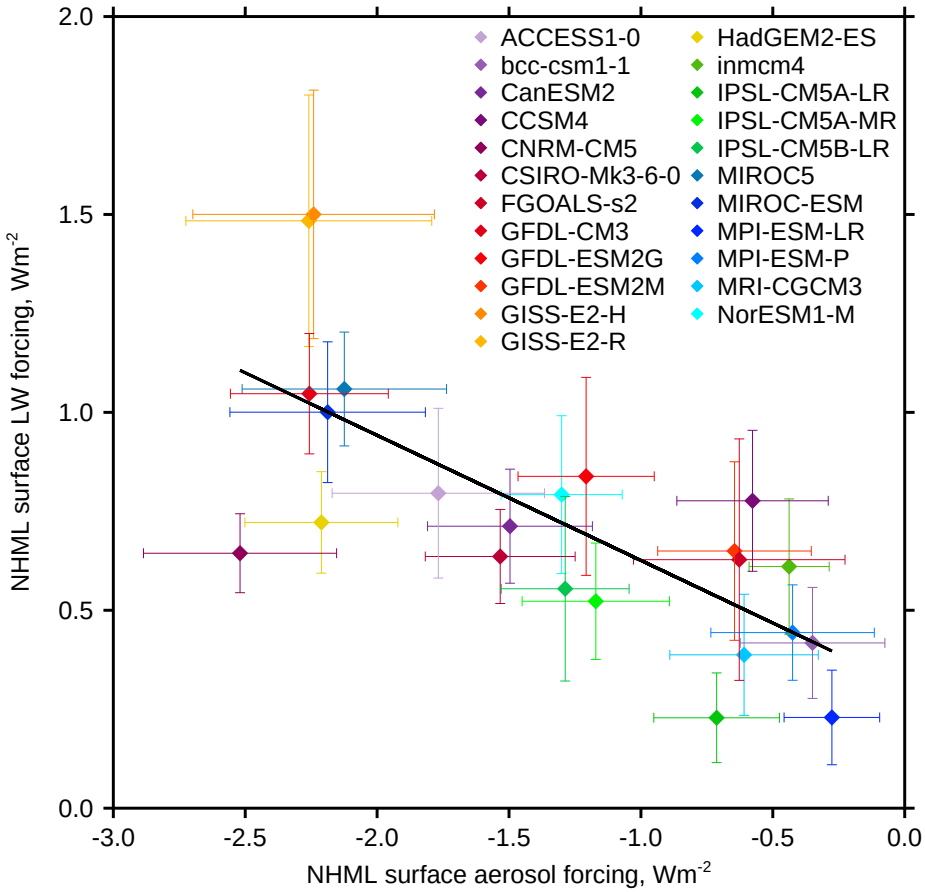

Figure 3.12: NHML surface LW forcing against NHML surface aerosol forcing for 23 CMIP5 models. NHML surface LW forcing is calculated using the technique outlined in Section 2.3.1.1, replacing surface SW fluxes with surface LW fluxes and taking the difference between mean NHML mean surface LW forcing before and after 1960. Horizontal and vertical error bars show an estimate of the 5–95% uncertainty ranges in NHML surface aerosol forcing and NHML surface LW forcing respectively.

3.4 Correcting NHML land precipitation observations

If the latest climate models do adequately simulate the twentieth century land precipitation response to aerosol (Zhang et al., 2007; Wu et al., 2013), an intriguing possibility is that trends in observed NHML land precipitation are unreliable. Significant precipitation gauge measurement biases have been previously identified, particularly in high-latitude regions (Groisman et al., 1991). Efforts to correct observations from recent decades for biases have resulted in increases in global land mean precipitation totals of 11.7% (Adam and Lettenmaier, 2003) and increases in winter higher mid-latitudes precipitation totals of 20–40% (Yang et al., 2005). For century scale observations time series inhomogeneities due to changes in instrumentation and recording practices can bias the true precipitation variability (Legates, 1995) and even change the sign of precipitation trends in some regions (Ding et al., 2007).

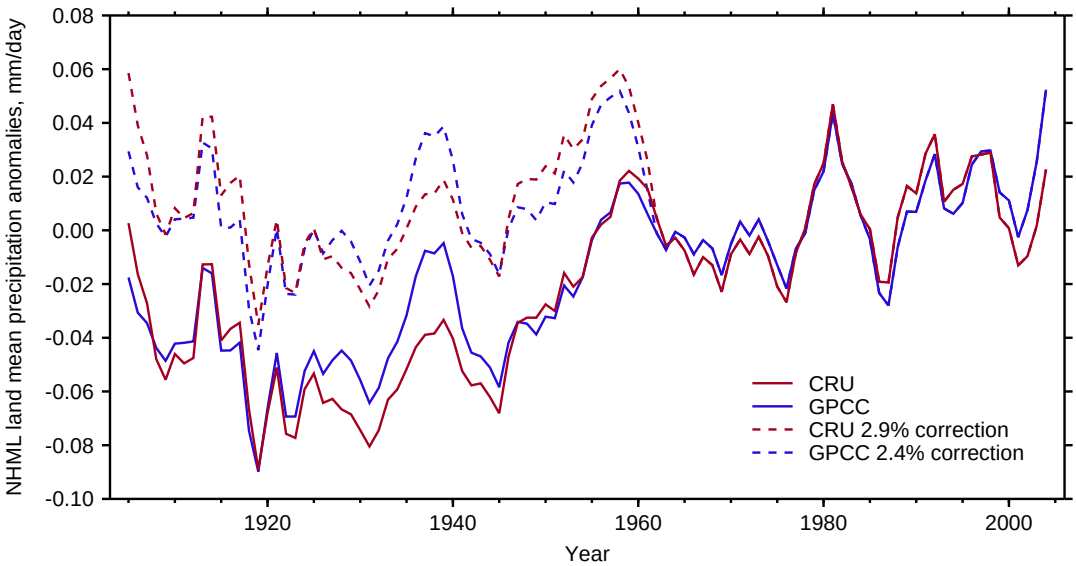

Figure 3.13: Five-year running mean NHML land mean precipitation anomalies for the CRU and GPCC datasets. Corrected data, where monthly total precipitation prior to 1960 is increased by 2.9% and 2.4% for CRU and GPCC respectively are shown (dashed lines). The NHML land precipitation offsets for the corrected data are consistent with IPCC equivalent NHML surface aerosol forcing and, in turn, the temperature gradient offsets and the NHT land precipitation offsets. All data are for 1905-2004 and anomalies are with respect to the period 1961-1990.

One region where numerous measurement errors have been documented is the

former Soviet Union (see Section 1.2.1.3). For example, in 1966 a wetting correction was introduced that increased typical individual station annual totals by 5–15% and in extreme cases by up to 30% for more northerly stations in winter (Groisman et al., 1991; Groisman and Rankova, 2001). However, even correcting each individual station in the former Soviet Union before inclusion in a gridded dataset would require extensive study of metadata, which may not exist (Groisman and Rankova, 2001), meaning that conventional NHML bias correction may not be possible. An alternative is to explore possible biases in the CRU and GPCC datasets using the physical framework that links aerosol, temperature and precipitation change. Monthly total precipitation values, and not just monthly precipitation anomalies, are available for both the CRU and GPCC datasets. An increase of 2.9% and 2.4% in monthly total precipitation before 1960 for CRU and GPCC respectively would generate NHML land precipitation offsets consistent with estimated aerosol forcing, the temperature gradient offset and the NHT land precipitation offset (Figure 3.13). Although this approach is crude, it serves to illustrate how sensitive trends in precipitation are to inhomogeneities of a magnitude frequently found in studies.

3.5 Conclusions

The effect of aerosols on regional temperature has been shown previously (Stott, 2003a; Shindell and Faluvegi, 2009). Here, a detectable mean temperature gradient response to aerosol forcing has been found, the size of which is dependent on the strength of aerosol forcing. Using this relationship, observed temperature is found to be in keeping with the estimated strength of real-world aerosol forcing. In agreement with existing literature, the NHT land mean precipitation response to remote extratropical aerosol forcing is found to be predictable across models, with observations again fitting into this framework. Theory suggests that aerosols should also have a significant impact on NHML land mean precipitation. Again, this is the case in CMIP5 models with stronger aerosol forcing associated with a

larger negative precipitation response. However, the observed NHML land mean precipitation response suggests that twentieth-century NHML aerosol forcing was positive, in disagreement with the modelled response and the scientific consensus on aerosol radiative effects.

The NHML region should contain the most valuable century-long observations. It is possible that model simulations of NHML land mean precipitation response to aerosol forcing are incorrect. Perhaps land surface processes such as soil moisture availability (Koster et al., 2004) are of large importance and are misrepresented by present models. However, there is evidence that present, widely used precipitation datasets may still contain significant biases, particularly in the early twentieth century, which restricts many precipitation studies to the second half of the twentieth century. It may be that, despite poorer spatial and temporal coverage, NHT land precipitation observations are more reliable through avoidance of measurement difficulties associated with high-latitude climate, such as snowfall undercatch. With scientific conclusions often based on agreement with observed datasets that are taken as truth (Groisman and Rankova, 2001) it is imperative that inhomogeneities in existing precipitation records are considered further. The NHML region is one of the most densely populated regions of the world. An improved understanding of past precipitation would help towards improved regional scale twenty-first-century projections, which are of huge value to policy-makers.

Chapter 4

Reconciling precipitation with runoff

In the last chapter it was shown that observed NHML land mean precipitation changes contain no evidence of a response to sharply increasing aerosol emissions in the mid twentieth century. GCMs, on the other hand, typically simulate a decrease in precipitation over this land region, with the magnitude of the decrease dependent on the strength of NHML surface aerosol forcing within the model. A possible explanation for this mismatch is that existing precipitation observations are not reliable, especially so in the early twentieth century. But is there any evidence for inhomogeneities?

One way of exploring this is to consider other components of the hydrological cycle. Precipitation and runoff are two intrinsically related components, but they are often studied in isolation (Xu et al., 2010). Consideration of observed runoff can complement observed precipitation and add extra weight to hypotheses for recent hydrological changes. Here, runoff observations are used to validate observed precipitation changes, with the aim of determining whether changes in precipitation are at odds with land surface hydrological change. Because the partitioning of precipitation into runoff and evapotranspiration is complicated and can be influenced by climate change, as well as direct human influences, care must be taken

in using runoff observations to test the reliability of precipitation observations. If there is evidence of major changes in how runoff changes with precipitation (the runoff-precipitation relationship) in the twentieth century, then this can be used as additional evidence – along with the apparent lack of a response to aerosol forcing – to question the reliability of observed NHML land mean precipitation.

4.1 The partitioning of precipitation at the land surface

In the absence of impacts of human activity and where natural changes in storage terms, such as soil infiltration and deep percolation, can be neglected – typically true on annual or longer timescales (Walter et al., 2004; Jung et al., 2013) – a simple water budget equation can be employed to a region consisting of one or more river catchments

$$P = Q + ET, \tag{4.1}$$

where P, Q and ET are precipitation, runoff and evapotranspiration respectively. Changes in runoff in many river catchments reflect changes in precipitation (Milliman et al., 2008). However, human activities can influence this water budget by changing the partitioning of precipitation into evapotranspiration and runoff and through changes in water storage, giving

$$P = Q + ET + \Delta S, \tag{4.2}$$

where ΔS is the net change in water storage due to human activities (Adam and Lettenmaier, 2008).

Dam and reservoir construction can affect seasonal water storage, but appear to have little effect at annual timescales (Adam et al., 2007; Adam and Lettenmaier, 2008). Melting ground ice in permafrost, due to human-induced warming (Bindoff et al., 2013), could lead to decreased water storage and has been invoked in

explaining the disparity between decreasing precipitation and increasing runoff in north Eurasian river catchments (Adam and Lettenmaier, 2008). Changes in water storage due to glacier and snow melt could also be significant. A human signal is clearly detectable in the global glacier mass loss during 1991 to 2010 and, to a lesser extent, over the longer industrial period (Marzeion et al., 2014). The greatest impacts of this loss in global glacier mass are expected in the local seasonality of water availability. Increased glacial melt acts to augment river catchment runoff. Extraction of groundwater for irrigation can enter the river system and increase runoff relative to non-irrigated conditions (Gerten et al., 2008). Such water storage changes are poorly understood, but here it is assumed that, away from the catchment scale, their influence on the water budget is minimal (i.e. $\Delta S \approx 0$).

4.1.1 Direct human influences

Humans can have a number of direct, but also poorly quantified, influences on the hydrological cycle, affecting the partitioning of precipitation into evapotranspiration and runoff. A major influence is water withdrawal for irrigation, which can have a big impact on runoff at the catchment scale (Milliman et al., 2008; Xu et al., 2010). A fraction of this withdrawal is not returned to runoff, defined as water consumption (Xu et al., 2010; Sterling et al., 2013). Although the significance of irrigation and other water withdrawals at the catchment scale is undisputed, there is less agreement over its influence on global mean runoff (e.g. Gerten et al. (2008); Sterling et al. (2013)). Some research suggests that decreases in global runoff resulting from irrigation are cancelled out by increases resulting from changes in land cover/land use (predominantly deforestation) (Gordon et al., 2005; Sterling et al., 2013).

4.1.2 Indirect human influences

Further to these direct influences of human activity are a number of indirect influences. Rising CO_2 can cause a transpiration-reducing plant physiological response and an accompanying increase in runoff (Gerten et al., 2008). Gedney et al. (2006) suggested this effect as the reason for increases in global runoff, as described by Labat et al. (2004), before contradictory subsequent work found no significant change in global runoff over the twentieth century (Milliman et al., 2008; Dai et al., 2009). Consideration of expanding vegetation is also required to quantify the overall CO_2 effect on global runoff in the twentieth century, the sign of which is, again, uncertain. Greenhouse gas-driven global warming (Bindoff et al., 2013) has caused a slight decrease in global runoff over the twentieth century, largely through higher summer evapotranspiration (Gerten et al., 2008), although this is dependent on sufficient moisture supply (Jung et al., 2010). Solar dimming associated with mid-twentieth-century aerosol emissions (Wild, 2012) has also been proposed as an indirect human influence on runoff, through forcing a decrease in evapotranspiration (Oliveira et al., 2011; Gedney et al., 2014). This effect is partly opposed by an increased diffuse fraction of solar radiation, which leads to increasing vegetation photosynthesis and transpiration (Huntingford et al., 2011). But the net aerosol impact is to increase runoff, with Gedney et al. (2014) finding a detectable increase in local river flow in response to increased NHML aerosol emissions in the 1960s and 1970s. This is of course a relative increase, identified through an idealised experimental setup – the difference between two land surface models both forced with twentieth-century observed meteorology, but with only one having a representation of aerosols – and an absolute decrease might still be expected should observed NHML land mean precipitation decrease as hypothesised in Chapter 3.

4.1.3 Approach

Despite the complicating mix of direct and indirect human influences on the hydrological cycle, the dominant driver of changes in runoff is changes in precipitation (Gerten et al., 2008; Milliman et al., 2008), which itself has been affected by human activities (Zhang et al., 2007; Wu et al., 2013, see Chapter 3). Gradual changes in the large scale runoff-precipitation relationship can be expected due to the more transient human influences, such as the CO_2 physiological effect. For example, increases in evapotranspiration on the North American continent have been documented (Milly and Dunne, 2001; Walter et al., 2004). But an approximately linear runoff-precipitation relationship is still expected, with sudden shifts in this relationship only at scales at, or close to, the catchment scale, perhaps due to deforestation or major dam construction. There appears to be no physical explanation for any such shifts at the continental scales investigated here, which would require, for example, a temporally-concentrated, continental-wide dam construction scheme.

Fekete et al. (2002) suggested that river discharge data, used to calculate catchment runoff, are more accurate than other water cycle components, with errors thought to be within 10–20%. This is an improvement on typical precipitation errors (Hagemann and Dümenil, 1998), with underestimates common due to wind-induced undercatch (see Section 1.2.1.2). Biemans et al. (2009) showed that the choice of gridded precipitation dataset can have a significant influence on the runoff simulated by hydrological models. Differences between datasets can be substantial at the catchment scale, especially in mountainous areas (Adam et al., 2006) and at high latitudes (Adam and Lettenmaier, 2003; Yang et al., 2005; Groisman and Rankova, 2001). The best agreement between datasets is found in areas of good spatial coverage of stations (often in the heavily-populated regions of North America and Europe), away from mountains and where the precipitation mix is dominated by rainfall. The difference between simulated and observed runoff can be tackled by tuning hydrological model parameterisations. But the

reliability of simulated runoff is inherently limited by the quality of observed precipitation input data (Biemans et al., 2009).

However, at the large scale, such as the NHML land mean, different precipitation datasets share more common variability and trends (see Section 3.1.2). Therefore, any major inhomogeneities that still exist must be common across all gridded precipitation datasets. Producing homogeneous records relies on quality observations of other meteorological variables (e.g. temperature, wind speed and precipitation type) and meta-information about individual stations (New et al., 2000; Groisman and Rankova, 2001) (see Section 1.2.1.3). These are both difficult to obtain, particularly in the early twentieth century. As such, traditional approaches to correcting observations may prove fruitless. The framework presented in Chapter 3 offers a basis for discovering inhomogeneities given the expected response to best estimates of real-world aerosol forcing. Yet there is still significant uncertainty in estimates of twentieth-century aerosol forcing (see Section 1.4.3) and a still-developing physical understanding of the aerosol effect on precipitation (see Section 1.4.4).

Using observed NHML mean runoff to assess observed NHML land mean precipitation relies on two key assumptions: 1) that runoff observations are immune to inhomogeneities of the type and magnitude common in precipitation observations, and 2) that in the absence of inhomogeneities the runoff-precipitation relationship is approximately linear. A statistical model is used to search for sudden shifts, defined as breakpoints, in the NHML land mean runoff-precipitation relationship. This approach is replicated at the continental scale to further pinpoint where any inhomogeneities in observed precipitation may exist. NHML land mean runoff is also analysed in six LSMs that form part of the TRENDY land surface model intercomparison project (Sitch et al., 2013). These models were all driven by observed NHML land mean precipitation, so offer the opportunity to search for breakpoints in the runoff-precipitation relationship that have arisen through a simulated physical process. Observed NHML land mean precipitation is then cor-

rected using its runoff counterpart and the implications for our understanding of the response of NHML land mean precipitation to mid-twentieth-century aerosol forcing discussed.

4.2 Data and methods

4.2.1 Observed NHML land mean runoff

Observed runoff is calculated using the Dai et al. (2009) continental discharge dataset. Discharge, or streamflow, is the temporally lagged, spatial integral of runoff over a river basin (Milly et al., 2005). It is possible to derive estimates of runoff for a catchment by dividing river discharge at a gauging station by the upstream catchment area (Fekete et al., 2002). Dai et al. (2009) produced a temporally complete streamflow dataset for 925 of the world's largest ocean-reaching rivers for 1948-2004, by infilling data gaps through considering nearby streamflow data, where possible. Any remaining data gaps were reconstructed by using a hydrological model and the correlation between precipitation and streamflow. Because the aim here is to contrast mid-twentieth-century aerosol-driven hydrological changes with overall twentieth-century changes, the temporally incomplete, observations-only, dataset that extends back to 1900 is used instead.

As in Chapter 3, the years 1905 to 2004 are considered. But here the water year, October-September, is used instead, with the first month contributing towards the period being October 1904. This minimises the lag effect associated with early-winter snowfall being locked within a catchment until the spring melt of the following calendar year (Dai et al., 2009). Owing to this lag effect, water year mean runoff for a given catchment is calculated only where all 12 months' worth of data are available. Further, the streamflow records are conditioned so that rivers must contain a minimum record length of 20 years in the 30-year 1961-1990 period (anomalies are calculated relative to this period). In Chapter 3, the

NHML region is defined as the latitude band between 30°N and 65°N. Here, it is stipulated that >80% of the catchment area upstream of the farthest downstream gauging station must fall within this band for a river to contribute to the NHML mean. This leaves a total of 148 rivers. Figure 4.1 shows the spatial coverage and record length of the 148 rivers.

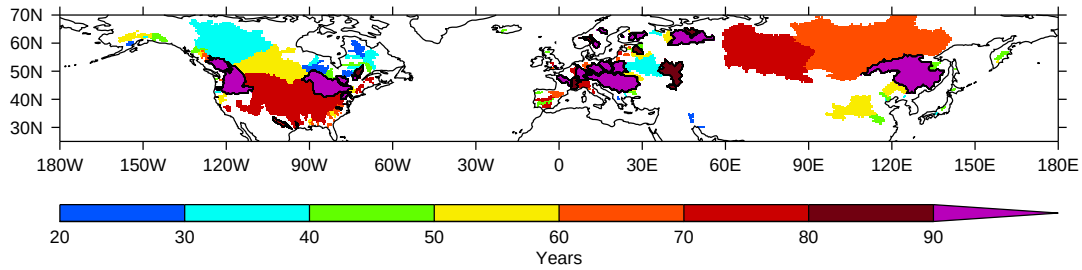

Figure 4.1: NHML catchment coverage of the 148 rivers that meet the conditions set out in Section 4.2, with the colour of the catchment representative of the (not always continuous) record length from the Dai et al. (2009) discharge dataset. River catchments with discharge records at least 80 years or longer are marked with black borders.

4.2.2 Observed NHML land mean precipitation

In Chapter 3, four gridded observational precipitation datasets were used in total. Spatially interpolated versions of two of the datasets, which offer complete global terrestrial coverage, were ignored. Here, such a product is necessary to allow for direct comparison with the spatial and temporal coverage of observed NHML mean runoff, with streamflow being a spatial integral of runoff over an entire catchment. To ensure an accurate comparison between precipitation and runoff in each catchment, high-resolution catchment masks are produced on a $0.5° \times 0.5°$ grid to match that of the precipitation dataset used. A version of the latest Climatic Research Unit high-resolution precipitation dataset (Harris et al., 2014), CRU TS3.21, is selected, which has been modified for use in driving the TRENDY models. This modification merged the monthly climatology of CRU TS3.21 with the National Centers for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR) reanalysis 1948-present dataset (Kalnay et al., 1996), bilinearly interpolated to the CRU TS3.21 resolution, to generate the diurnal and daily variability required to drive the TRENDY models

at 6-hourly time-steps (see Fisher et al. (2013) for further details). The merged dataset (CRUNCEP) offers a finer temporal resolution for modelling purposes, but monthly precipitation totals remain the same. Figure 4.2 shows the twentieth century NHML annual spatial coverage for both runoff and the precipitation (with and without the time varying runoff mask applied) station network. During some later decades runoff observations offer roughly a five-fold increase in spatial coverage over precipitation. This is of course still an overestimate for precipitation, since gauges only provide a point measurement and many grid boxes are poorly gauged, especially in mountainous regions where precipitation is highly variable (Adam et al., 2006).

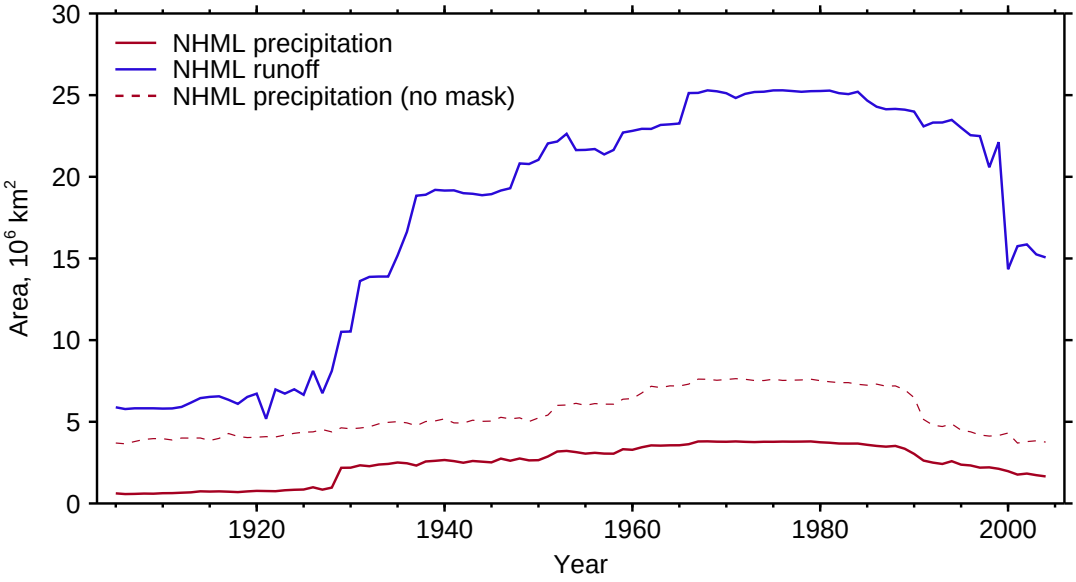

Figure 4.2: Maximum annual NHML land area coverage of observed runoff and precipitation using the Dai et al. (2009) discharge dataset and the CRU TS3.21 precipitation dataset (using the raw product, which only considers grid boxes where real observations exist) respectively. Both river catchment masks and the CRU TS3.21 dataset are computed on a high resolution 0.5° × 0.5° grid. Also shown (dashed line) is the maximum coverage of the CRU TS3.21 precipitation dataset (also non-spatially interpolated) without the time varying runoff mask applied.

4.2.3 TRENDY models

An ensemble of six LSMs are used, summarised in Table 4.1. The models are forced over the historical period with CRUNCEP precipitation, as well as other observed climate variables that are derived from merging CRU and NCEP data, and changing CO_2 concentrations (Sitch et al., 2013; Le Quéré et al., 2014). These

Model name	Abbreviation	Original resolution	Reference
Community Land Model 4.5	CLM4.5	1.25° × 0.9375°	Oleson et al. (2013)
Joint UK Land Environment Simulator	JULES	1.875° × 1.25°	Clark et al. (2011); Best et al. (2011)
Lund-Potsdam-Jena	LPJ	0.5° × 0.5°	Sitch et al. (2003)
LPJ-GUESS	LPJG	0.5° × 0.5°	Smith et al. (2001); Ahlström et al. (2012)
LPX-Bern	LPX	1.0° × 1.0°	Stocker et al. (2013)
ORCHIDEE-CN	OCN	1.0° × 1.0°	Zaehle and Friend (2010); Zaehle et al. (2010)

Table 4.1: Overview of the ensemble of six TRENDY LSMs.

runs were also driven by historical land use changes, calculated from the HYDE dataset (Klein Goldewijk and Verburg, 2013). Although Yang et al. (2015) report some sensitivity of runoff trend in TRENDY models to land use change at the catchment scale, conclusions here are insensitive to the inclusion/exclusion of the historical land use dataset. Simulated runoff is available as output at a monthly frequency. Model output not already at 0.5° × 0.5° resolution is regridded using bilinear interpolation. Simulated runoff is then masked to the varying spatial and temporal coverage of the runoff (streamflow) observations.

4.2.4 CMIP5 models

Observed NHML land mean precipitation is compared with that modelled by 15 CMIP5 models. In contrast to Chapter 3, GCMs with a historical (all forcings) experiment ensemble size of just one are omitted (Table 2.1), as well as the two GISS GCMs (for the reasons detailed in Section 3.3.2). Simulated precipitation in the CMIP5 models is also regridded to 0.5° × 0.5° resolution and masked to the availability of runoff observations. The NHML land mean precipitation offset is calculated as before (using equation (2.9), described in Section 2.3.1.2). The NHML surface aerosol forcing estimates for CMIP5 models are unchanged from Chapter 3, using the technique outlined in Section 2.3.1.1. The estimate of IPCC equivalent NHML surface aerosol forcing is again used for observed precipitation (see Section 3.3.1). Figure 4.3a shows the updated version of Figure 3.9c. Overall, there is little difference between the plots, with the stark contrast between the observed precipitation response and the modelled precipitation responses still evident. Removing this mask, and instead considering all grid boxes in the CRUNCEP dataset between 30°N and 65°N, an even more robust relation-

ship between the CMIP5 precipitation response to aerosol forcing and the CMIP5 NHML surface aerosol forcing is seen (Figure 4.3b). This suggests that masking the precipitation dataset to the runoff observations weakens this robust signal, but it remains significant.

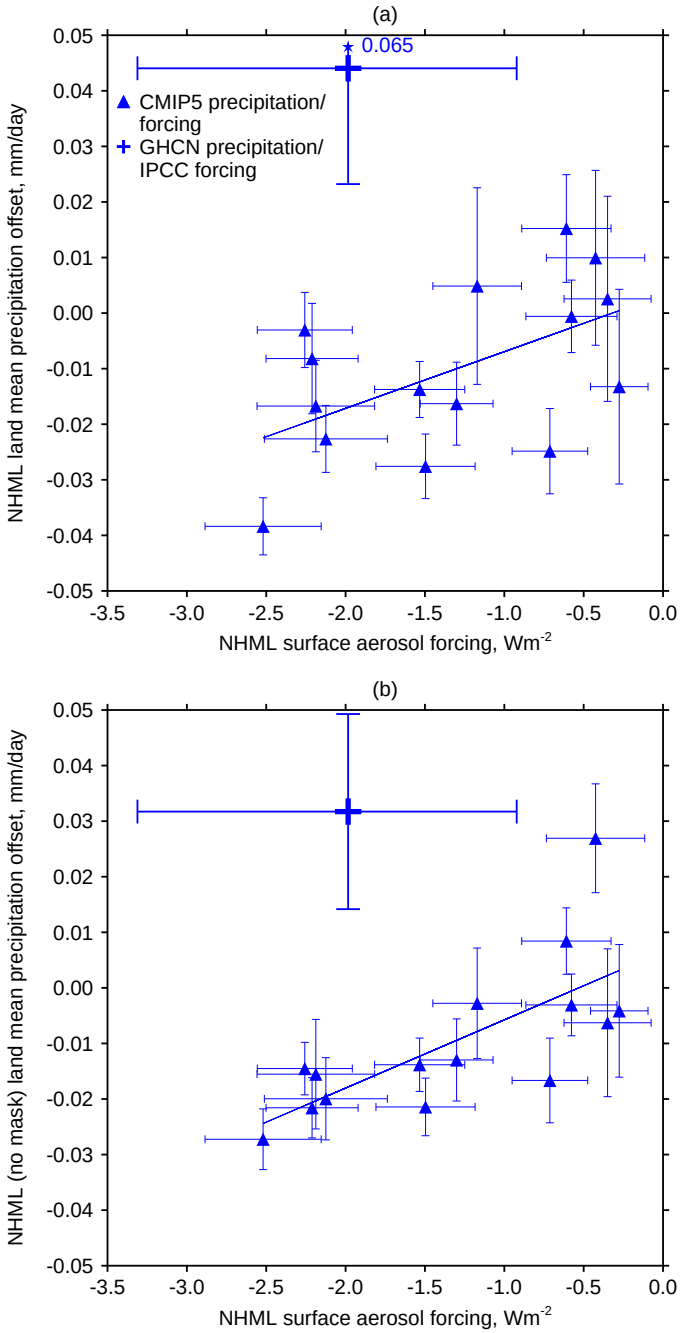

Figure 4.3: NHML land mean precipitation offset against NHML surface aerosol forcing. Error bars represent the 5–95% uncertainty range (see Sections 2.3.1.1 and 2.3.1.2), with IPCC equivalent NHML surface aerosol forcing and the associated 5–95% uncertainty range used for the observed data point (see Section 3.3.1). (a) CRUNCEP masked to the availability of runoff observations, and (b) all CRUNCEP grid boxes between 30°N and 65°N. The value next to the error bar in (a) indicates where the uncertainty extends beyond the scale. There is a significant positive correlation between precipitation offset and aerosol forcing across the 15 CMIP5 GCMs in both (a) and (b) ($r = 0.54$, $p < 0.05$ and $r = 0.87$, $p < 0.01$ respectively).

4.3 The runoff-precipitation relationship

Ideally, Figure 3.9c would be reproduced using just NHML land mean runoff, with an expectation that interannual, decadal and multidecadal changes in runoff should match those in precipitation. However, the representation of runoff in CMIP5 models is found to be suboptimal, with the runoff-precipitation relationship frequently non-linear. Bias in the simulated precipitation climatology leads to bias in aridity and, in turn, the runoff coefficient (runoff/precipitation), in line with the Budyko framework (Budyko, 1974). Within this framework the partitioning of precipitation into runoff and evapotranspiration at annual mean catchment scales varies as a function of aridity (Greve et al., 2015). Aridity, or the dryness index, is the dimensionless ratio of potential evapotranspiration (PET) to precipitation. The evaporative index, the dimensionless ratio of evapotranspiration to precipitation, is then dependent on the dryness index. The relationship is described by the non-linear Budyko curve (or a derivation of the Budyko curve), which is constrained by two physical limits; the atmospheric demand for water ($ET < PET$; the red dashed 1:1 line in Figure 4.4) and the atmospheric supply of water ($ET < P$; the blue dashed horizontal line in Figure 4.4).

The implications of a bias in the simulated precipitation climatology is demonstrated in Figure 4.4 for the Yellow river catchment in China. The CMIP5 models typically simulate too much precipitation over the catchment. In this situation the aridity is incorrectly simulated as energy limited ($PET/P < 1$), rather than moisture limited ($PET/P > 1$). Consequently the sensitivity of changes in runoff to changes in precipitation is too high. The fit to observed data is made using the Fu equation (Fu, 1981; Zhang et al., 2004), a parametric form of the Budyko curve. This one-parameter function relates the evaporative index to the dryness index such that

$$\frac{ET}{P} = 1 + \frac{PET}{P} - \left[1 + \left(\frac{PET}{P}\right)^{\omega}\right]^{1/\omega}, \tag{4.3}$$

where ω is an empirical parameter that is calibrated against local data.

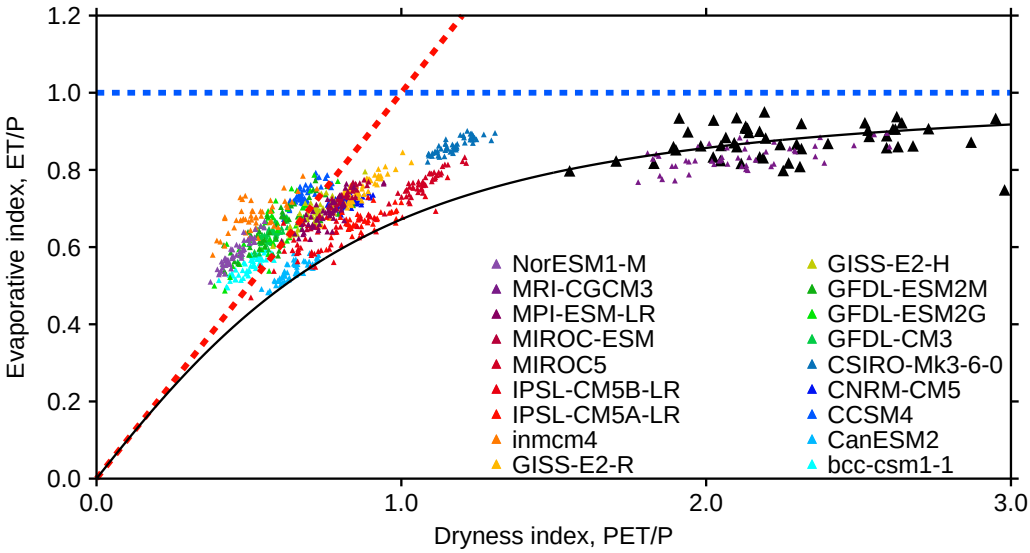

Figure 4.4: Evaporative index (ET/P) against dryness index (PET/P) for the Yellow river catchment. Annual mean data (1951-1999) are shown for observations (black triangles) and multiple CMIP5 models (coloured triangles). Note that a subset of 18 of the 23 CMIP5 models listed in Table 2.1 is shown (due to availability of data). The curve is Fu's equation (equation (4.3)) with the basin specific ω calculated by minimising the least squares errors between the estimated and the observed runoff ($\omega = 2.44$). The red dashed 1:1 line represents the atmospheric demand for water ($ET < PET$) and the blue dashed horizontal line represents the atmospheric supply of water ($ET < P$).

Runoff outputs from GCMs have also been described as biased because of the representation of runoff processes (Hirabayashi et al., 2013). This may be a consequence of low resolution or insufficient subgrid parameterisation (Weiland et al., 2012). In some GCMs runoff may be simply advected straight to the ocean, meaning that runoff is poorly represented in larger river catchments because time lags between runoff generation and return to the ocean are not considered (Falloon et al., 2011; Weiland et al., 2012). Because of the coarse resolution and reduced complexity of land surface schemes in GCMs relative to offline LSMs, changes in runoff are often investigated using direct GCM runoff output integrated to river discharges using a river routing scheme (Milly et al., 2005; Hirabayashi et al., 2013), or through using observed precipitation or GCM precipitation output to drive LSMs (Gerten et al., 2008; Schewe et al., 2014), as has been done with the TRENDY models.

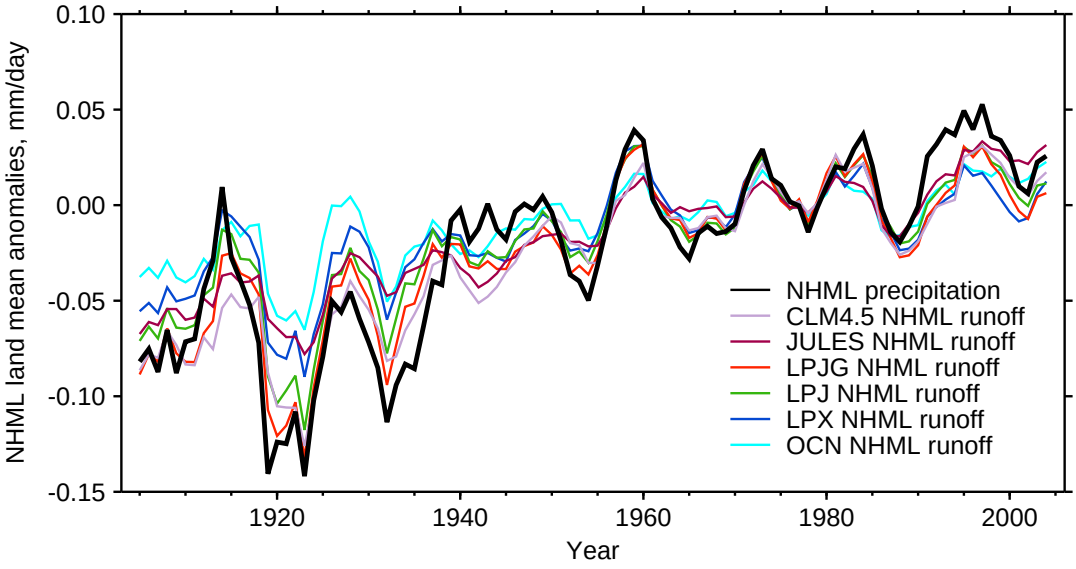

Figure 4.5: Five-year running mean NHML land mean model simulated runoff and observed precipitation anomalies. Data are from the six TRENDY LSMs (coloured lines) and CRUNCEP dataset (thick black line) for runoff and precipitation respectively.

Dataset	Reference	Trend (mm/year/century)
Precipitation (from Chapter 3)		
CRU	Harris et al. (2014)	**30.5 ± 9.1**
GPCC	Becker et al. (2013)	**28.8 ± 8.3**
GHCN	Vose et al. (1992)	**32.8 ± 9.6**
Zhang	Zhang et al. (2007)	**45.4 ± 10.8**
Precipitation		
CRUNCEP	Fisher et al. (2013)	**44.8 ± 12.6**
Runoff		
Dai	Dai et al. (2009)	3.2 ± 8.5

Table 4.2: Trends in observed NHML land mean precipitation and runoff, with precipitation trends shown for the four datasets used in Chapter 3 (observations only and calendar year), as well as the CRUNCEP dataset used in the current study (spatially interpolated and water year). Errors are the 5–95% confidence intervals, with values in boldface significantly different from zero ($p < 0.05$).

Model/ *Dataset*	Trend (mm/year/century)
CLM	**40.5 ± 7.5**
JULES	**32.5 ± 4.5**
LPJG	**37.4 ± 10.2**
LPJ	**32.1 ± 9.1**
LPX	**22.4 ± 8.1**
OCN	**19.0 ± 6.1**
TRENDY mean	**30.6 ± 7.2**
CRUNCEP	**44.8 ± 12.6**

Table 4.3: Trends in NHML land mean runoff in the TRENDY LSMs and TRENDY multimodel mean, as well as the NHML land mean precipitation trend in the observed CRUNCEP dataset used to drive the LSMs. Errors are the 5–95% confidence intervals, with values in boldface significantly different from zero ($p < 0.05$).

4.3.1 Contrasting simulated and observed runoff

Although direct runoff output is considered, rather than routed discharge, the six TRENDY LSMs all simulate NHML land mean runoff that is strongly correlated with observed NHML land mean precipitation used to drive the experiments (Figure 4.5). The twentieth century observed CRUNCEP precipitation trend is found to be positive and significant, consistent with the four precipitation datasets used in Chapter 3 (Table 4.2), which use only actual observations (no spatial interpolation) at a coarser resolution and are not constrained to the availability of runoff observations. Trends in NHML land mean precipitation are robust to the choice of dataset (some of which favour fewer long-term homogenised records over more short-term records), infilling of data through interpolation and spatial coverage. Interestingly, the runoff trend is positive and significant in all six models, but in each case less than the trend in observed CRUNCEP precipitation (Table 4.3). With the increase in precipitation not being matched by an increase in runoff it would appear that the LSMs are all able to capture the already mentioned, well documented, increase in evapotranspiration (Milly and Dunne, 2001; Walter et al., 2004).

Observed NHML land mean runoff shows a near-zero trend over the twentieth

century (Table 4.2), in line with what is expected given mid-twentieth-century aerosol forcing. Annual mean runoff is strongly correlated with annual mean precipitation ($r = 0.62$, $p < 0.01$). However, inspection of the NHML land mean runoff and precipitation time series appears to show a split in the relationship around the 1930s or 1940s (Figure 4.6). Simply adjusting the pre-1960 precipitation, as in Chapter 3, to the pre-1960 runoff, using the linear regression fit of post-1960 five-year mean NHML land mean runoff against five-year NHML land mean precipitation, the positive trend in precipitation is reduced and becomes non-significant. Figure 4.7 shows five-year mean NHML land mean runoff anomalies against five-year mean NHML land mean precipitation anomalies, split into a pre-1940 group and a post-1940 group (here, 1940 has been arbitrarily selected as a breakpoint in the runoff-precipitation relationship). On first impression there would appear to be a clear breakpoint, with two lines with similar slopes but different intercepts.

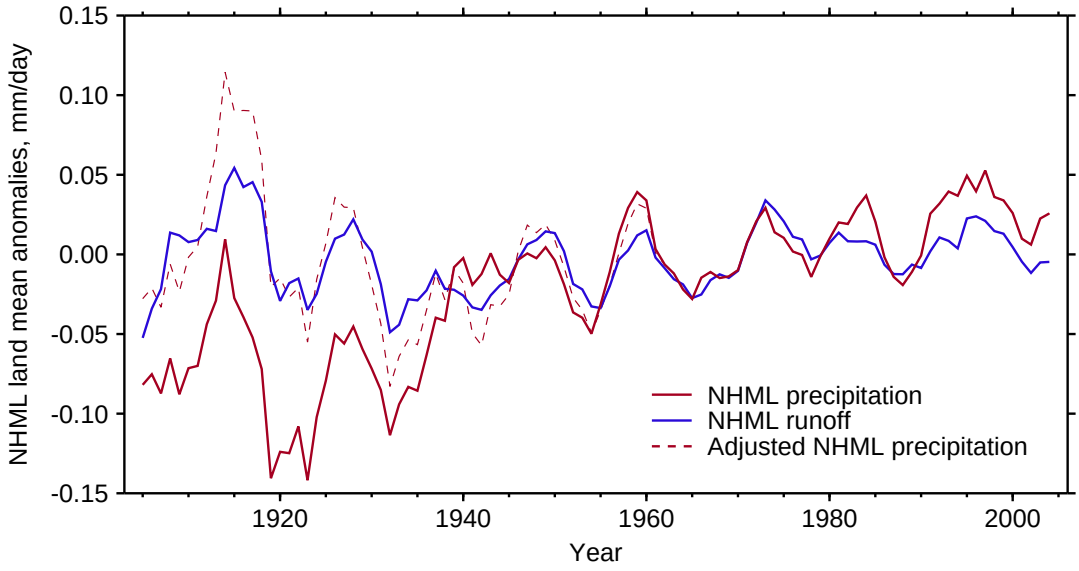

Figure 4.6: Five-year running mean NHML land mean runoff and precipitation anomalies. Data are from the Dai et al. (2009) discharge dataset and CRUNCEP dataset for runoff and precipitation respectively. The adjusted NHML land mean runoff time series is constructed using pre-1960 runoff values and the post-1960 five-year mean runoff-precipitation linear regression fit. All data are for the water year, October-September, 1905-2004 and anomalies are with respect to the period 1961-1990.

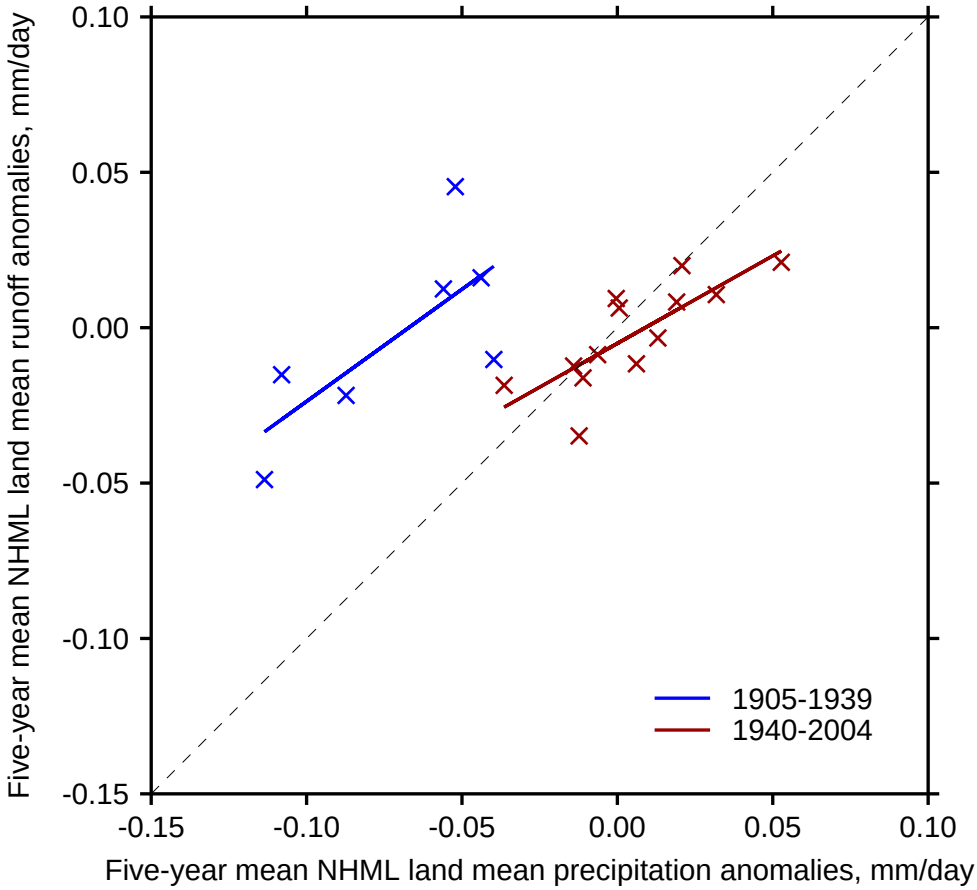

Figure 4.7: Five-year mean NHML land mean runoff-precipitation relationship, split into a pre-1940 group and a post-1940 group.

4.3.2 A breakpoint detection model

Determining whether a breakpoint is a real feature of this relationship requires a statistical test. Testing for breakpoints, also referred to as inhomogeneities (Peterson et al., 1998), change points (Reeves et al., 2007) and step changes (McCabe and Wolock, 2002), in hydrometeorological variables is a long established practice. Many different breakpoint detection methods are available and they vary in the number of breakpoints they can detect, the assumptions made on data distribution and whether the timing of the breakpoint is known before testing (see Reeves et al. (2007) and Ferguson and Villarini (2012) for more in depth discussions). The approach here differs from many of these studies in that a method is required to detect breakpoints in the relationship between two intrinsically linked hydrometeorological variables, rather than the time series of just one variable.

Breakpoint detection begins by fitting a model with four parameters, representing

two lines with unique slopes and intercepts

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 z_i + \beta_3 z_i P_i + \epsilon_i \tag{4.4}$$

where Q are the five-year mean NHML land mean runoff anomalies, P are the five-year mean NHML land mean precipitation anomalies and z is a binary variable coded 0 for data points in the first time segment and coded 1 for data points in the second time segment. The β coefficients are parameters to be estimated. The residuals ϵ are assumed to be identically and independently distributed with zero mean. By varying the length of the two time segments it is possible to test for breakpoints at a number of five-year intervals. The minimum length of each time segment, or group, is set at 25 years (a sample size of five). Therefore, breakpoints are searched for between 1930 and 1980 inclusive. An analysis of variance (ANOVA) F-test is used to determine whether equation (4.4), at any given division, offers a significantly better fit to the data than a simple one slope, one intercept model

$$Q_i = \beta_0 + \beta_1 P_i + \epsilon_i. \tag{4.5}$$

However, because multiple tests are performed – 11 tests in total at the $p = 0.05$ level – then there is a probability greater than 0.05 of finding at least one breakpoint. Because there is a probability greater than 0.05 that evidence against the null hypothesis (that equation (4.5) is the best fit) appears by chance, the significance level at which each of the 11 tests is performed requires adjustment to reduce the probability of a type I error back to 0.05.

To investigate this problem further, pseudo data is simulated, creating 10,000 datasets of runoff and precipitation anomalies. Random data for both precipitation and runoff are generated with a normal distribution with mean 0 and variance 1 (the variance is assumed to be constant between time segments). The probability of finding at least one breakpoint is 0.213. Performing each of the 11 tests at the $p = 0.0078$ level gives a probability of finding at least one breakpoint of 0.05. Therefore, if the F-statistic at any of the 11 intervals is significant at the 0.78%

level, the more complex model (equation (4.4)) is fitted. If β_3 in equation (4.4), known as the interaction parameter, is not significant ($p > 0.05$) it is assumed that β_3=0 and a more parsimonious, additive, model is fitted

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 z_i + \epsilon_i \tag{4.6}$$

used for separate regression lines with the same slopes. However, some change in slope might be expected due to the more transient, indirect human influences discussed that could be prevalent at such large scales (e.g. Walter et al. (2004); Gedney et al. (2014)). To avoid including the effect of slope changes in the breakpoint estimate, β_2 from equation (4.6) is taken to be the breakpoint magnitude. The difference in β_2 values from equation (4.4) and equation (4.6) is due to slope changes between the two time segments. This residual is typically small and the conclusions in this Chapter are not dependent on the choice of either equation (4.4) or equation (4.6). Although there could be more than one inhomogeneity in the observed precipitation and runoff records, the assumption is made that only one major breakpoint exists in the runoff-precipitation relationship. Further, this approach requires the assumption that any major inhomogeneity occurs on a timescale of years (Groisman and Rankova, 2001) rather than decades. Where possible breakpoints are identified at more than one time division, the breakpoint is chosen as the model fit that minimises the residual sum of squares (RSS).

4.3.3 Breakpoint detection results

Potential negative breakpoints in the observed NHML land mean runoff-precipitation relationship are identified at three time divisions – 1930, 1935 and 1940 – with the model fit to the 1930 breakpoint found to minimize the RSS (Table 4.4). Because this is the earliest interval at which a breakpoint can be detected using five-year means, the analysis is repeated using three-year means. Using five-year means is common when analysing changes/trends in precipitation and in detection and attribution studies (Min et al., 2011; Wu et al., 2013) as it

Equation	β_0	β_1	β_2	β_3
Equation (4.4)	**0.063**	**0.794**	**-0.066**	-0.364
Equation (4.6)	**0.040**	**0.475**	**-0.044**	—

Table 4.4: The estimated β coefficients from a model fit to a 1930 breakpoint for both equation (4.4) and equation (4.6). Values in boldface are significant at the $p = 0.05$ level.

reduces the signal to noise and avoids problems with autocorrelation. The autocorrelation function for both the observed runoff and precipitation decays towards zero at a time lag of three years (not shown). Repeating the analysis for three-year means – offering a detection range between 1920 and 1990 – gives near-identical results with the breakpoint identified in 1932 (accounting for the multiple tests problem). This suggests the result is robust. Because of the highly variable nature of precipitation, at smaller scales – continental scale and below – only five-year means are considered in further analyses to improve the signal to noise.

With a breakpoint detected the approach is extended to determine where geographically the breakpoint occurs. While downscaling to the catchment scale is limited by temporal coverage and poor signal to noise, downscaling to the continental scale should offer more insight (Figure 4.8). The continental scale analysis points towards North America as the likely source of the breakpoint (Figure 4.8a). A North American breakpoint is detected in 1935. The agreement between observed Europe land mean runoff and precipitation is excellent and the relationship is free of a breakpoint (Figure 4.8b). The Asia land mean runoff and precipitation time series, which include all Russian river catchments (Figure 4.8c), suggest that a breakpoint might exist there in the first 30-40 years. Although one is not detected, reducing the minimum time segment length to 20 years, instead of 25 years, detects a breakpoint in 1925. The first $\sim$25 years of the Asia time series are dominated by the Amur river, a catchment with particularly poor precipitation station spatial coverage at that time. This will undoubtedly contribute towards the 1930 breakpoint found in the NHML land mean runoff-precipitation relationship. The role of the Amur river and poor spatial coverage of the early twentieth century rain gauge network is addressed in Section 4.4.2. Beyond about 1930

the agreement between precipitation and runoff in Asia is reasonable. However, a contrast between decreasing precipitation and increasing runoff from the 1950s to the 1980s is evident, as found in other studies (Adam and Lettenmaier, 2008; Milliman et al., 2008).

The breakpoint model is also applied to the TRENDY LSM NHML land mean runoff-precipitation relationships. Breakpoints are detected in two models, CLM4.5 and JULES, both in 1950. However, the breakpoints are of opposite sign (positive) to the observed breakpoint (negative) and are about half the size. Also, using the multimodel mean NHML land mean runoff no breakpoint is detected. The CLM4.5 and JULES simulations here have the lowest resolution of the six TRENDY models. Previous work has shown that a model performs much better in simulating the hydrological balance of various European regions with increased resolution (Elguindi et al., 2011). Investigating the differences between observed and simulated runoff spatially, the North American continent again stands out. Differences between observed and multimodel mean twentieth-century runoff change reflect the discrepancy over North America (Figure 4.9). This is expected given that a North American breakpoint has been detected in the 1930s. There are notable differences for individual rivers worldwide, probably reflecting runoff changes that are not modelled. But these differences largely cancel out across each of the European and Asian continents. Over North America, changes in modelled runoff are evidently greater (more positive) than changes in observed runoff.

4.3.4 Adjusting observed precipitation

Given the 1930 breakpoint in the observed NHML land mean runoff-precipitation relationship, what is the effect of adjusting the precipitation time series to bring it in line with runoff? An adjustment is applied that preserves the more subtle changes in the runoff-precipitation relationship, so that the slopes of the two lines in equation (4.4) are unchanged. Pre-1930 precipitation is instead adjusted

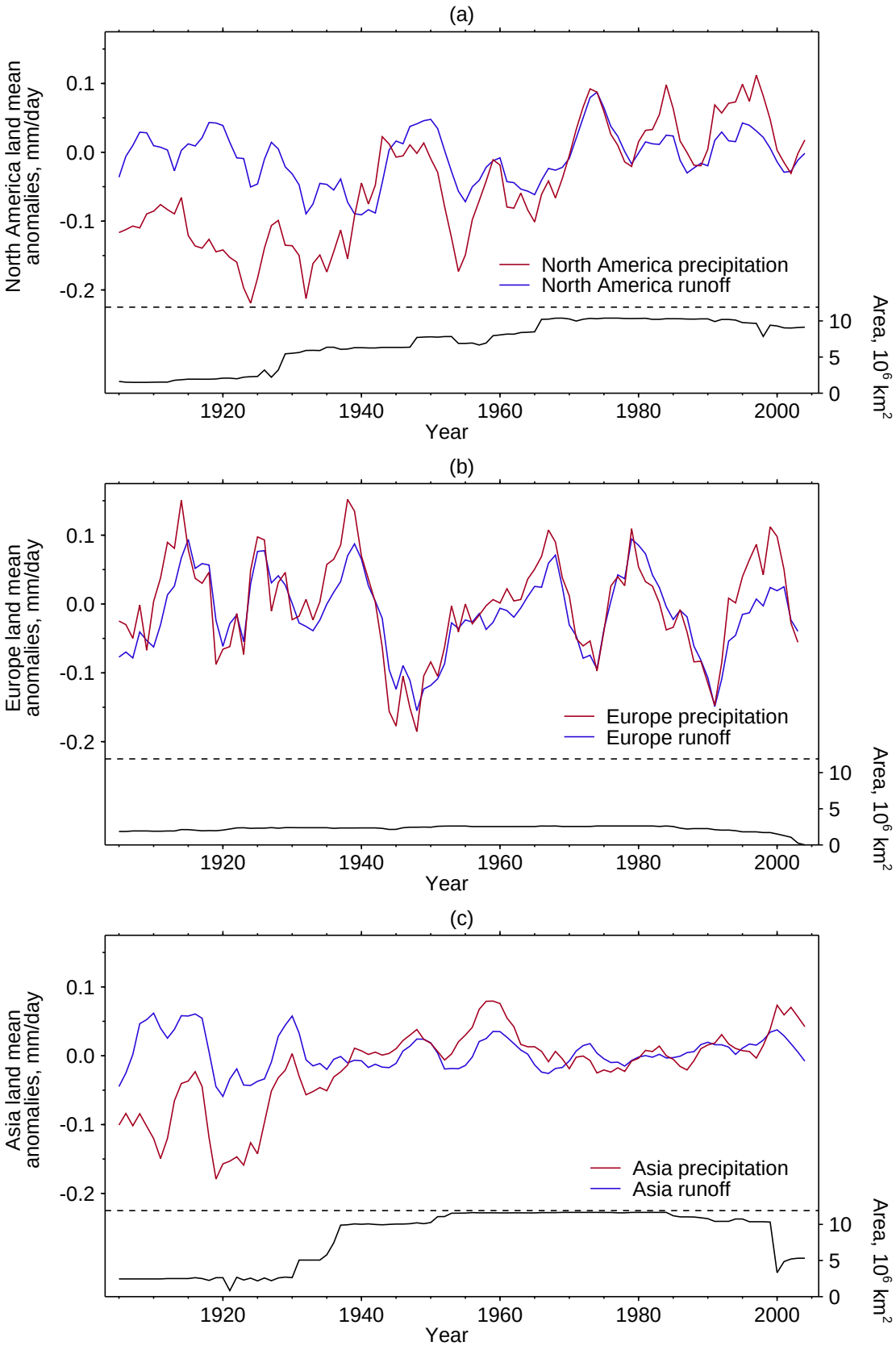

Figure 4.8: Five-year running mean land mean runoff and precipitation anomalies and annual land area coverage of observed runoff. (a) North America, (b) Europe, and (c) Asia.

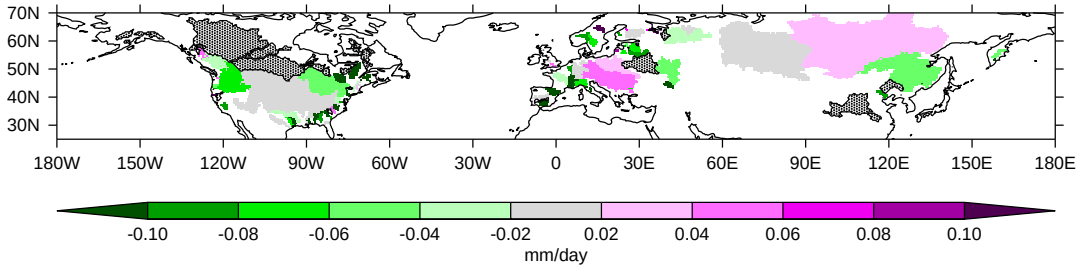

Figure 4.9: Difference in observed and TRENDY multimodel mean twentieth-century runoff change. Twentieth-century runoff change is the difference between mean runoff before and after 1960, for years with observed data, and is calculated at the catchment scale. A minimum of 20 years' worth of data are required from both the pre-1960 and post-1960 periods – the latter a necessary requirement to calculate a climatology for a given river catchment – so that only a subset (92) of the 148 river catchments are shown here. Stippling represents the remaining 56 rivers that do not meet this condition.

(increased) so that $\beta_2 = 0$ in equation (4.6). The effect of this is to reduce the positive trend in twentieth-century precipitation from 44.8 ± 12.6 mm/year/century to 7.2 ± 13.6 mm/year/century (Figure 4.10). Reproducing Figure 4.3 with adjusted observed precipitation, the observed NHML land mean precipitation offset of 0.002 ± 0.021 mm day^{-1} is equivalent to an aerosol forcing of -0.1 ± 2.0 Wm^{-2}, calculated using the NHML land mean precipitation offset-aerosol forcing regression coefficients (Figure 4.11). Prior to this precipitation adjustment an offset of 0.044 ± 0.021 mm day^{-1} gave an estimate of aerosol forcing of 4.0 ± 2.0 Wm^{-2}.

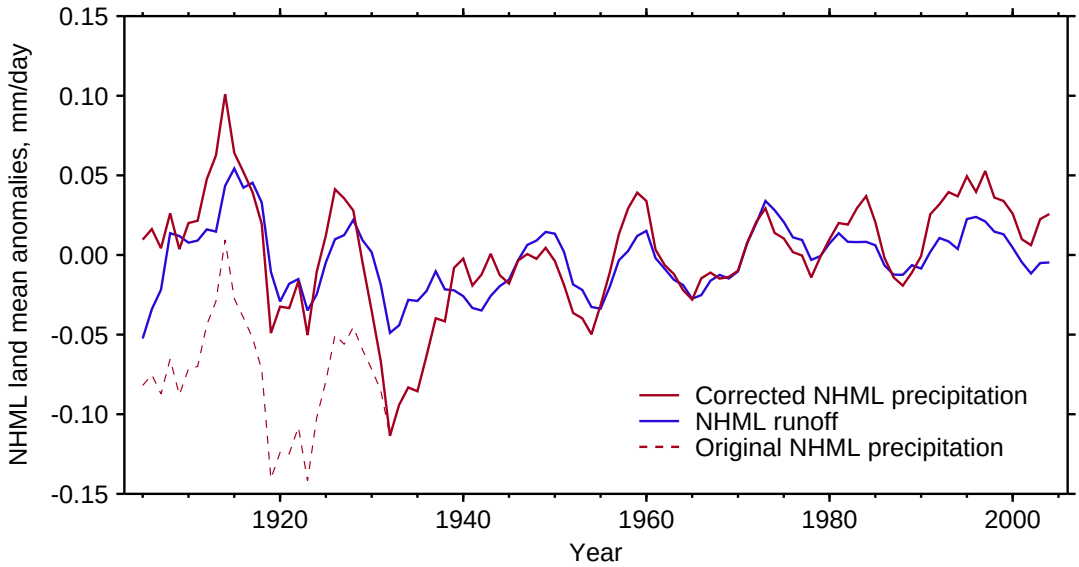

Figure 4.10: Five-year running mean NHML land mean runoff and precipitation anomalies. This is an updated version of Figure 4.6, showing the adjusted precipitation record, given a 1930 breakpoint in the runoff-precipitation relationship. The original precipitation record is also shown for reference.

This adjusted precipitation gives an offset consistent with negative aerosol forc-

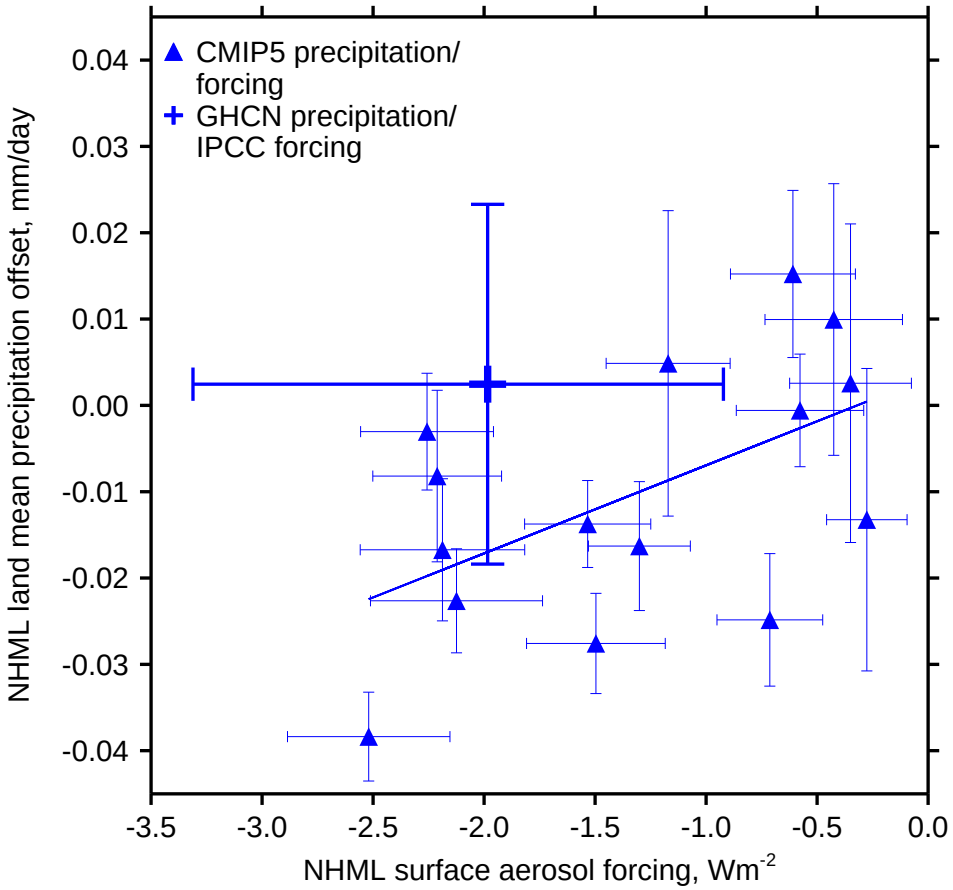

Figure 4.11: NHML land mean precipitation offset against NHML surface aerosol forcing. This is an updated version of Figure 4.3, but the adjusted precipitation record, given a 1930 breakpoint in the runoff-precipitation relationship, is used to calculate the observed NHML land mean precipitation offset.

ing. The difference between the best estimate of real-world NHML surface aerosol forcing of -2.0 [-3.3 to -0.9] Wm^{-2} (see Section 3.3.1) and that suggested by the precipitation offset has reduced from 6.0 Wm^{-2} to 1.9 Wm^{-2}. Quantitatively similar results are found when using the simple technique outlined in Section 4.3.1 (Figure 4.6), where the precipitation time series is adjusted, given the post-1960 runoff-precipitation relationship. This reduces the discrepancy from 6.0 Wm^{-2} to 1.7 Wm^{-2}. These results still somewhat contradict similar analyses performed for observed Northern Hemisphere tropical land mean precipitation and observed temperature gradient (between the NHML region and Southern Hemisphere extratropical region), which both show an aerosol response in agreement with expectations (see Section 3.3).

4.4 Explaining the breakpoint

4.4.1 Eliminating possibilities

Twentieth-century, surface-based meteorological datasets will always contain inhomogeneities, owing to their changing spatial coverage in time (see Section 1.2.1.4). These inhomogeneities and biases may be shared between different gridded precipitation datasets, as different groups use many common records. From Figures 4.1, 4.2 and 4.8 it can be seen that the annual NHML land area coverage increases around the time that the breakpoint is found. This is when the three great, Arctic Ocean draining Eurasian rivers (Ob, Yenisey and Lena) and the Mississippi river begin contributing to NHML mean runoff anomalies. A near-constant spatial mask between 1905 and 2004 can be considered by selecting only river catchments with discharge records at least 80 years or longer (catchments with black borders in Figure 4.1). Because CRU TS3.21 is spatially interpolated the true spatial coverage of the underlying station network will still vary somewhat. This approach greatly reduces the spatial coverage but the breakpoint remains a robust feature (Figure 4.12), although 1935 is now the most likely timing. Changing spatial coverage in time could possibly lead to a breakpoint as different climatic regions are incorporated into the NHML mean. This is because a certain change in precipitation should not be expected to produce the same change in runoff in an arid region as in a humid region (see Section 4.3). However, the TRENDY LSMs realistically represent the range in the runoff coefficient across different climates, but do not contain breakpoints of the sign or magnitude of the observed breakpoint.

In Section 4.1 the impact of dam and reservoir construction on the NHML mean runoff time series was dismissed. At the catchment level the influence of flow regulation (the ratio of total reservoir storage capacity to mean annual discharge) could be important. Vörösmarty et al. (1997) showed that 20–30% of global mean annual discharge was now stored in reservoirs constructed in the late twentieth century. Water impoundments can increase evaporative losses and therefore de-

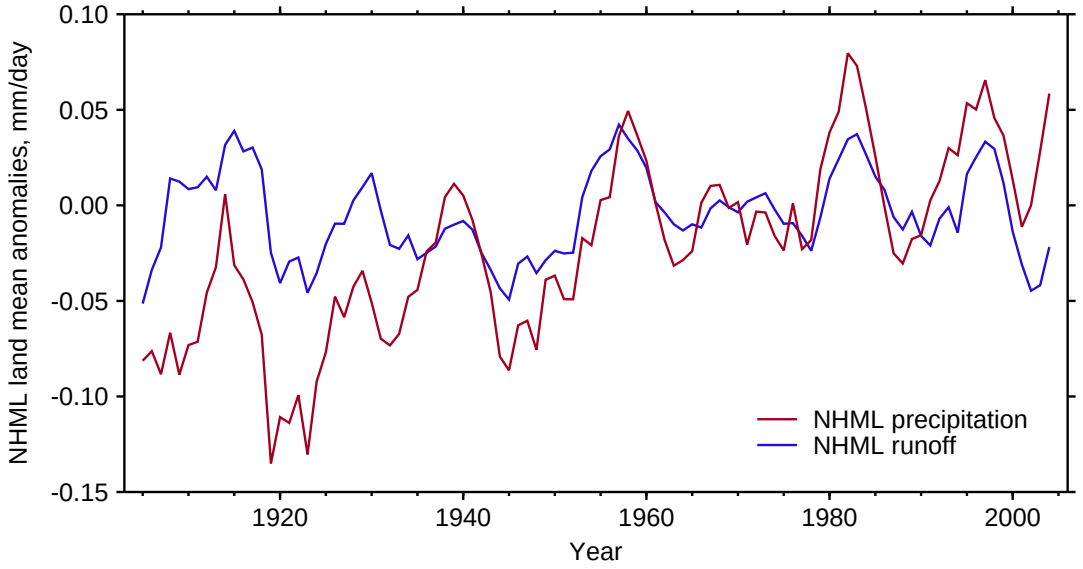

Figure 4.12: Five-year running mean NHML land mean runoff and precipitation anomalies using a near-constant spatial mask (river catchments with discharge records at least 80 years or longer (see Figure 4.1)). Data are from the Dai et al. (2009) discharge dataset and CRUNCEP dataset for runoff and precipitation respectively. All data are for the water year, October-September, 1905-2004 and anomalies are with respect to the period 1961-1990.

crease runoff. Evaporative losses from reservoirs have been shown to have an affect on individual catchments, mostly in arid regions with low streamflow (e.g. Vörösmarty and Moore III (1991)), but appear to be insignificant at continental scales (Vörösmarty et al., 1997; McClelland et al., 2004). Schwalm et al. (2014) recently found that including a representation of water management structures in terrestrial biosphere models made little difference to simulated runoff over the United States. Of course, the filling of a large reservoir following the construction of a dam causes an abrupt decrease in catchment runoff. This could lead to an anomalous annual mean runoff value. But following filling, streamflow should return to pre-dam levels so such impoundments are unlikely to cause breakpoints.

Dam and reservoir construction allows for easier access to water resources and typically leads to more water withdrawal, often for irrigation. Water withdrawn from reservoirs for irrigation can cause substantial decreases in runoff due to enhanced evapotranspiration. Irrigation is likely to account for a significant fraction of the stark runoff decreases observed in the Yellow river catchment in China since the 1960s (Xu et al., 2010). The Yellow river is one of the 148 that met the conditions set out in Section 4.2.1. To determine what effect highly irrigated

catchments have on the above findings, irrigation index values (the proportion of area equipped for irrigation (from Siebert et al. (2005)) to naturalised discharge (discharge prior to direct human influences)) are taken from the Nilsson et al. (2005) study. Values are available for 42 out of 52 large rivers considered here with an upstream catchment area of greater than 50,000 km^2. These 42 large rivers account for 88% of the total area of all 148 river catchments. As found by Milliman et al. (2008), irrigation and flow regulation indices (the latter are also taken from the Nilsson et al. (2005) study) are correlated (Figure 4.13). A total of 8 rivers (Sacramento, Dnepr, Yellow, Don, Ebro, Colorado, Tagus, Liao) are found to exceed both an irrigation index of 200 and a flow regulation of 20, defined by Milliman et al. (2008) as thresholds above which deficit rivers (rivers where changes in runoff are considerably less than changes in precipitation) typically fall. Removing these rivers and considering only the remaining 34 (Figure 4.14), a significant breakpoint is still found (a 1935 breakpoint provides a marginally better fit to the data than a 1930 breakpoint).

4.4.2 Proposing possibilities

A subject worthy of consideration is the quality of the interpolated CRUNCEP precipitation observations. While complete spatial coverage is desirable, it is important to consider the true spatial coverage of the underlying station network. The area represented per station within a catchment, defined as the area per station (APS), is much greater across the Asian continent than the North American and European continents (Table 4.5), particularly so in the 1905-1929 period. However, as already discussed (see Section 4.3.3), this is also a time period when spatial coverage of runoff observations across the Asian continent was largely limited to the Amur river, the fifth largest river of the 148 in terms of catchment area and ranking fifth in terms of 1905-2004 mean annual streamflow. There are no available streamflow observations for the three great Arctic Ocean draining Eurasian rivers (Ob, Yenisey and Lena) between 1905 and 1929, when their APS

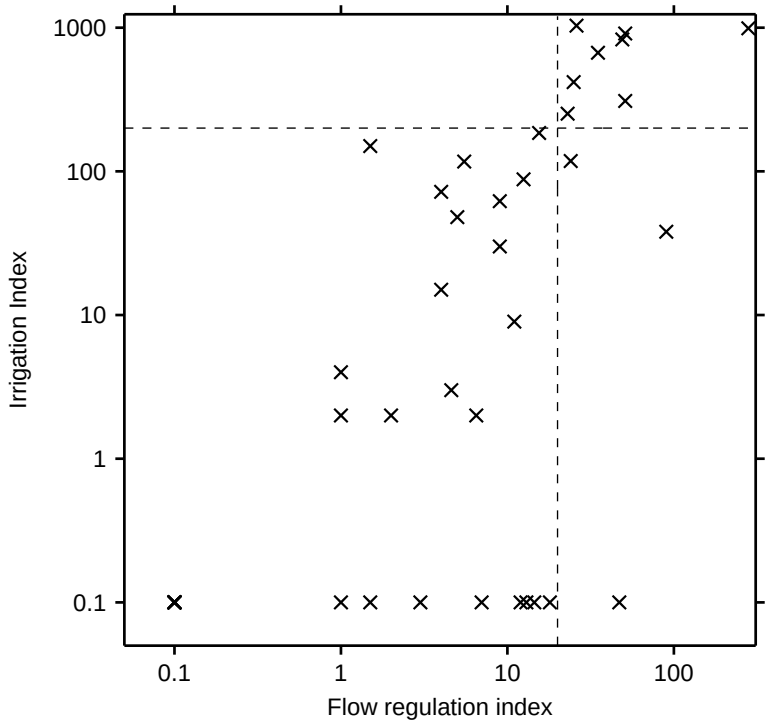

Figure 4.13: Flow regulation indices (the ratio of total reservoir storage capacity to mean annual discharge) and irrigation indices (the ratio of area equipped for irrigation to naturalised discharge) for 42 large rivers with with an upstream catchment area of greater than 50,000 km^2. Zero values are plotted as 0.1 to highlight those catchments with no irrigation potential but some level of flow regulation. The horizontal and vertical dashed lines mark the irrigation index threshold of 200 and the flow regulation index threshold of 20 respectively. Milliman et al. (2008) showed that deficit rivers typically fall above these thresholds.

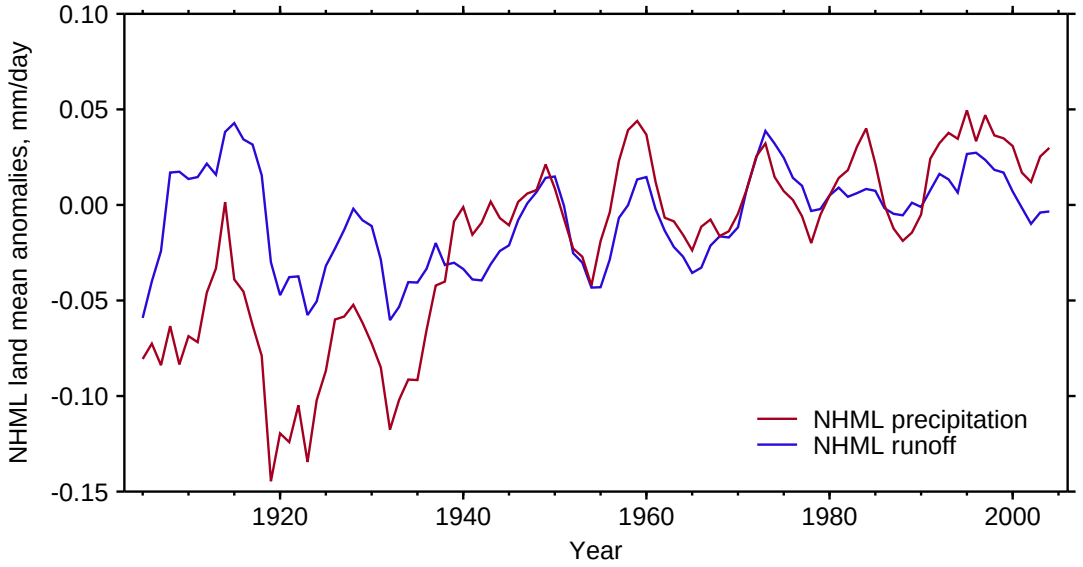

Figure 4.14: Five-year running mean NHML land mean runoff and precipitation anomalies using 34 large river catchments with an area of greater than 50,000 km^2 and an irrigation index of less than 200 or a flow regulation index of less than 20 (see text). Data are from the Dai et al. (2009) discharge dataset and CRUNCEP dataset for runoff and precipitation respectively. All data are for the water year, October-September, 1905-2004 and anomalies are with respect to the period 1961-1990.

	North America	Europe	Asia
Variable mask			
1905-1929	14,866	31,677	155,338
1930-2004	18,474	16,384	36,038
Fixed mask			
1905-1929	24,107	27,918	245,153
1930-2004	20,790	17,455	42,151

Table 4.5: The APS of the North American, European and Asian continents for the 1905-1929 and 1930-2004 periods (before and after the NHML land mean runoff-precipitation breakpoint is detected). Values are shown for the variable mask, where precipitation is masked to the availability of runoff observations (the spatial mask varies in time), and the fixed mask, where precipitation is masked to the catchments of all 148 rivers (the spatial mask is constant in time). Units are given in km^2.

values were extremely high. As such, Asia mean APS values are noticeably different in this period depending on whether a fixed temporal mask of all rivers is used or whether precipitation is masked to the availability of runoff observations. When masked to the availability of runoff observations the Amur catchment has the worst mean annual APS of all rivers. The poor APS of the Amur river means that Asia land mean precipitation is likely to be more vulnerable to inhomogeneities in the pre-1930 period.

Although the APS might explain the Asian continent contribution towards the 1930 NHML land mean runoff-precipitation breakpoint, it has less merit in explaining the more prominent 1935 North American breakpoint. Up until 1935 the St Lawrence and Columbia river catchments account for much of the spatial coverage of North American runoff observations. The St Lawrence and Columbia are the fourth and fifth largest river catchments of the 72 in North America respectively, with the former draining into the Atlantic and the latter draining into the Pacific. Combined aggregated Pacific and Atlantic draining river catchments contain a runoff-precipitation breakpoint in 1935, although the β_1 and β_2 coefficients in equation (4.4) are not significant ($p = 0.25$ and $p = 0.29$ respectively). No breakpoint is found when considering the aggregated Gulf of Mexico draining rivers (dominated by the Mississippi catchment). While attributing the exact location and cause behind a breakpoint in the runoff-precipitation relationship is beyond the scope of this thesis, this does help narrow the focus of future work

into precipitation inhomogeneities.

The 1930s saw a notable hydroclimatic event across North America in the form of the "Dust Bowl", a period characterised by low precipitation amounts and high temperatures (e.g. Schubert et al. (2004)). This precipitation deficit period is seemingly captured by the North America precipitation time series (Figure 4.8a) and in a hypothetical time series of observed North America land mean precipitation (following the methodology of Section 4.3.4) (not shown). The LSMs – driven by observed precipitation – do not appear to simulate this breakpoint. A breakpoint is detected in one model (LPX) in 1940, which is of the same sign as the observed breakpoint but roughly half the magnitude. Interestingly, this is no longer detected when considering only river catchments with discharge records at least 80 years or longer (as above), implying some sensitivity to spatial coverage. Overall, the LSMs do not support a real change in the runoff-precipitation relationship over the North American continent around this time.

4.5 Conclusions

In Chapter 3 observed NHML land mean precipitation was found to be missing an expected response (Wu et al., 2013) to well-documented mid-twentieth-century aerosol forcing (Lamarque et al., 2010; Wild, 2012). This is despite a robust signal across CMIP5 GCMs, with the magnitude of the response dependent on the strength of aerosol forcing (see Section 3.3). It was suggested that inhomogeneities in the precipitation record could be part of the reason for this disagreement (see Section 3.4). Here, precipitation observations have been reconciled with complementary runoff observations, which represent a key land surface component of the hydrological cycle.

Variations in interannual, decadal and multidecadal runoff are known to be predominantly forced by variations in precipitation (Dai et al., 2009). Occasionally, long-term trends in runoff are shown to differ from long-term trends in precipita-

tion (Milliman et al., 2008). This is particularly prevalent at the catchment scale, where direct human influences, such as water withdrawals for irrigation (Milliman et al., 2008; Xu et al., 2010) and land cover/land use change (Piao et al., 2007; Sterling et al., 2013) can have significant impacts. At larger scales, differences in runoff and precipitation trends can be attributed to changes in evapotranspiration (Milly and Dunne, 2001; Walter et al., 2004), perhaps because of increasing temperatures due to rising greenhouse gas concentrations (Huntington, 2008; Krakauer and Fung, 2008) and augmented, at least locally, by aerosol-induced dimming (Gedney et al., 2014). Consideration of the plant physiological response to rising CO_2 levels is also necessary (Gedney et al., 2006). However, while these influences may have an effect on the runoff-precipitation relationship in the NHML land mean, they are not expected to cause breakpoints.

Testing for breakpoints in the NHML land mean runoff-precipitation relationship, three potential candidates at five-year intervals were found (1930, 1935 and 1940), with a negative breakpoint in 1930 found to provide the best fit to the data. Runoff anomalies pre-1940 are noticeably greater than precipitation anomalies. This lends support to the hypothesis suggested in Chapter 3 that pre-1960 precipitation anomalies should be greater. It was found that an increase of 2.9% in monthly total precipitation before 1960 in the CRU TS3.21 dataset would generate a negative NHML land mean precipitation offset consistent with estimated aerosol forcing (see Section 3.4). Results from this chapter do not support an increase in precipitation totals across the entire pre-1960 period or the magnitude of the increase suggested. However, adjusting the precipitation record before the 1930 breakpoint (using the runoff record) does account for some of the discrepancy between the (lack of) observed precipitation aerosol response and that suggested given the most recent estimates of real-world aerosol forcing (Myhre et al., 2013). Importantly, the range of aerosol forcing suggested by this adjusted response is within the real world aerosol forcing range.

Using six LSMs as part of the TRENDY intercomparison project (Sitch et al.,

2013) it is possible to show that this breakpoint is unlikely to be a consequence of known physical processes. Driven by observed precipitation, four of six models show no breakpoint, whereas two show small, positive breakpoints in 1950. Although not nearly as apparent as the breakpoint in the observed runoff-precipitation relationship, and perhaps just a consequence of the choice of statistical model, this warrants further attention. However, it is concluded that the modelled runoff-precipitation relationship is free of breakpoints of the sign and magnitude found in observations. Perhaps the most obvious shortcoming in this technique of using observed precipitation to drive LSMs is that land surface feedbacks on the atmosphere are neglected, because of the absence of coupling (Harding et al., 2014; Schewe et al., 2014), limiting the potential accuracy of results. Future efforts to incorporate more complex land surface schemes into GCMs should aid understanding of the importance of such interactions.

A continental scale search for breakpoints found that the North American continent was the likely source for the observed NHML-wide breakpoint. Changes in measurement practice and gauge type are just two of many explanations for such an inhomogeneity in precipitation records (see Section 1.2.1.2). Runoff records are assumed to be less susceptible to inhomogeneities (Hagemann and Dümenil, 1998), but some will undoubtedly exist and should also be investigated in much the same way as precipitation records. In addition, it would be unwise to wholly discount the influence of dams, with dam construction in the Columbia river catchment, for example, beginning in the 1930s, notably with the Bonneville Dam. Although most work suggests that the influence of dams on annual mean discharge at large scales is minimal (Vörösmarty et al., 1997; McClelland et al., 2004). The St Lawrence and Columbia river catchments deserve particular attention as their runoff and precipitation observations contribute significantly to the pre-1935 North American land means.

Further, there are likely to be some shortcomings in the methodology. Firstly, not all precipitation inhomogeneities can be expected to appear as an obvious

discontinuity, with some biases likely to emerge slowly over time, perhaps because of changes in vegetation near to a rain gauge (Legates, 1995). Also, just one statistical model is used to test for breakpoints. It would be useful to see how robust these findings are to different techniques, while also considering the runoff and precipitation records independently. Problems associated with using spatially interpolated data have also been highlighted. The Amur river catchment was identified as a region with a sparse precipitation station network in the early twentieth century. Therefore, a direct comparison between runoff and precipitation here would contrast the complete spatial coverage of runoff (inherent in streamflow measurements) with just a few point measurements of precipitation.

It is not possible to provide a robust explanation for a 1930s breakpoint. One of the more widely documented systematic changes in precipitation measurement occurred in the former USSR. In 1966-1967 a wetting correction was introduced (see Section 1.2.1.3). But no such systematic changes are documented in the 1930s. Instead, it might be the case that the spatial coverage of early-twentieth-century observations, combined with the under-representation of, say, mountainous regions leaves NHML land mean precipitation observations prone to inhomogeneities during this period. Without targeted studies, however, this remains speculative. Some possibilities behind such a breakpoint have been eliminated (Section 4.4.1), including the influence of dams and reservoirs, as well as irrigation. This is to be expected since the 1930s preceded major direct human influences on the land surface component of the hydrological cycle.

Although changes in runoff closely track changes in precipitation the relationship between these two variables is complicated. For example, recent research has shown that a precipitation shift from snow to rain may lead to a decrease in runoff (Berghuijs et al., 2014) and this may play a role in explaining the steady evapotranspiration increases seen over the NHML region and particularly the North American continent. Additionally, it should be noted that NHML land mean runoff alone, although showing a near-zero twentieth-century trend, does

not show an obvious decrease around the 1960s when aerosol emissions increased. However, recent work using multiple idealised experiments and different driving observations in JULES has shown that increased aerosol emissions and associated solar dimming increased Northern Hemisphere river flow around this time (Gedney et al., 2014). Correcting for this effect would further reduce the discrepancy between the observed and predicted precipitation response to aerosol forcing. This result neglects the importance of land surface feedbacks on precipitation, but further illustrates the complicated nature of attributing changes in runoff. Regardless, NHML land mean runoff observations suggest NHML land mean precipitation observations to be in error, as surmised in Chapter 3. But the discrepancy between observed and modelled precipitation over the NHML land region cannot be wholly accounted for. So, while inhomogeneities are likely to exist in precipitation records, it is concluded that model simulated precipitation response to aerosol forcing should still be questioned.

Chapter 5

Conclusions

This chapter presents the conclusions. The key findings are listed, together with some of the limitations of the work and directions for future research. Whether the initial aims and objectives have been met is discussed, as well as the contribution of the work to the research field.

5.1 Key findings

A key focus for this thesis was changes in Northern Hemisphere land precipitation, largely because model simulated output here can be compared and contrasted with century-long observations. But to begin, Chapter 2 considers global mean precipitation change, because this is the scale at which changes are best understood. The first aim of this research (see Section 1.7) was "to determine the impact of contrasting direct effects of different climate forcings, particularly GHGs and aerosols, on twentieth-century global mean precipitation change". Even though global mean precipitation change has been studied in depth, it has largely been studied in a way that is difficult to relate to twentieth-century climate. Here, information is used about the different direct effects of key forcing agents on the troposphere to present an idealised global mean precipitation-temperature relationship. Using just the two key forcings, GHGs and sulphate aerosols – modelled

as increasing linearly and increasing as a mid-twentieth-century step change respectively – the relationship should appear as two GHG-driven straight lines offset by the effect of this step change in sulphate aerosol forcing.

This offset is a clearly identifiable feature of the global mean precipitation-temperature relationship in CMIP5 models, but would seem to be larger in some models than others. As theorised, the size of this precipitation offset is shown to be strongly correlated with the strength of global aerosol forcing. This is true in CMIP5 models and the older CMIP3 models. In this idealised framework it is shown that the magnitude of the global mean precipitation offset is a function of a coefficient, which varies little between models, and the step change in sulphate aerosol forcing. Although the real world does not evolve in the same way as considered in this simple model, the strength of aerosol forcing would appear to explain a lot of the variance in the global mean precipitation offset.

A natural consequence of downscaling to look at regional scale precipitation change is that features that can be robustly detected and attributed at the global scale become less obvious. The climate system is controlled by different processes at these scales. However, Chapter 3 presents evidence that an aerosol signature should be expected in NHML land mean precipitation and NHT land mean precipitation (as well as the NHML-SHEXT temperature gradient). This is seen in model simulations, with the size of the NHML and NHT land mean precipitation offset again significantly correlated (although less strongly than in the global mean) with local and remote NHML surface aerosol forcing respectively (such a correlation exists for the mean temperature gradient too). But while the observed mean temperature gradient and observed NHT land mean precipitation offsets agree well with what is predicted given estimates of real-world NHML surface aerosol forcing and the relationship across CMIP5 models, the observed NHML land mean precipitation offset is not even of the expected sign. This chapter satisfactorily answers the research aim "to explore the extent to which a similar aerosol signature might be expected to be seen and is seen in regional land mean

precipitation changes in both models and observations".

It does not, however, offer any insight into the reason behind these contrasting signals in models and observations. The fact that NHML land precipitation changes show little evidence of an aerosol response may be a result of inhomogeneities in the precipitation record. Alternatively, it could be a consequence of shared inadequacies in the latest GCMs. Chapter 4 attempts to explore the reason by tackling a third research objective "to determine whether observed regional precipitation changes are consistent with observed changes in regional runoff". Inconsistency would suggest that serious questions should be raised about the suitability of the whole record of twentieth-century precipitation observations in the NHML region for ongoing scientific research.

Strong evidence is found to further question the reliability of these key precipitation observations, particularly in the first half of the twentieth century. A clear breakpoint in the runoff-precipitation relationship is discovered in the 1930s. This is obvious just from NHML land mean precipitation and runoff time series, which share common interannual, decadal and multidecadal variability. But as a consequence of this breakpoint the whole twentieth-century trends are noticeably different. Although the partitioning of precipitation into runoff and evapotranspiration is complicated and can be influenced by humans in many ways, multiple lines of evidence are presented to suggest that the runoff-precipitation relationship should be linear and free of such breakpoints. Perhaps the most useful line of evidence comes from an ensemble of six LSMs that were all driven with the same observed precipitation dataset. These suggest that a real increase in evapotranspiration over the twentieth century was likely, but are free of any breakpoints of the sign and magnitude of those found in the observed relationship.

This strongly suggests that NHML land mean precipitation observations, our longest and most comprehensive precipitation observations, contain spurious trends that do not show evidence of a response to aerosol forcing and are not in keeping with complementary runoff observations. Rather inconclusively, adjusting

precipitation given runoff does not fully account for the discrepancy between the observed precipitation offset and that expected given the estimate of actual NHML surface aerosol forcing. But given the various uncertainties this is perhaps to be expected. So, rather frustratingly it is concluded that this discrepancy between observed and modelled precipitation is most likely due to inhomogeneities, but the accuracy of model simulated precipitation responses to key climate forcings – both homogeneous and heterogeneous – should still be questioned. Hopefully, the emergence of satellite-derived precipitation observations will, in time, provide a more suitable means for testing the response of increasingly complex climate models.

5.2 Limitations

There are several limitations to the approach used in Chapter 2. One is the methodology used for diagnosing radiative forcing. The linear forcing-feedback regression analysis to derive forcing time series (see Section 2.2.1) could be inappropriate if Earth's energy imbalance is non-linear with global mean surface air temperature in a fixed forcing scenario (i.e. the abrupt4×CO_2 experiments) so that the climate feedback parameter, α, is time-varying. Evidence for non-linearity was found by Andrews et al. (2015) who showed that the climate system becomes more sensitive to climate forcings in later years of these idealised experiments due to a dependence on the pattern of SST change. The assumption of constant climate sensitivity in a transient forcing situation is likely to be even less valid (Andrews et al., 2015). But this method for diagnosing forcings remains a very useful tool for explaining the spread in behaviour across climate models (Forster et al., 2013).

Another limitation is the use of SW forcing to represent aerosol forcing. An improved method would see the net forcing calculated for anthropogenic aerosol forcings only simulations for all CMIP5 models. But these experiments were listed

as low priority within the CMIP5 framework and were therefore only performed by a small number of modelling centres. Using available simulations, however, it is shown that historical all forcings SW forcing is a good proxy for historical anthropogenic aerosol forcings only net forcing (see Section 2.3.1.1).

Of course, in the absence of further simulations it is difficult to separate the influences of sulphate aerosol forcing and black carbon aerosol forcing. This is problematic because the simple approach presented in Chapter 2 to show the idealised precipitation-temperature relationship ignores the effect of black carbon forcing, even though it has been shown in previous studies to have a significant direct effect (if black carbon aerosol forcing was assumed to be directly proportional to sulphate aerosol forcing the magnitude of the precipitation offset would be expected to increase). There is no easy way to overcome this. To counteract the potential problem that arises from the fact that TOA black carbon forcing can be positive but surface black carbon forcing negative, SW/aerosol forcing is diagnosed at the surface as well as the TOA. This ensures a consistent sign convention – a more negative surface aerosol forcing, be it from black carbon aerosol or sulphate aerosol – leads to a more negative precipitation offset. It materialises that, even regionally, the relationship between the precipitation offset and the aerosol forcing is not hugely sensitive to the choice of a TOA or surface definition.

While it is expected that sulphate and black carbon constitute the dominant aerosol forcings, other aerosols could also be important (see Figure 1.1). Twentieth-century nitrate and organic carbon radiative forcing has also been negative and, while small, this cumulative forcing is clearly non-negligible. The representation (or lack of) these additional aerosol species and their associated forcing in GCMs could explain some mismatches between simulated and observed precipitation and temperature, especially regionally. It was noted, for example, that an overestimate of Southern Hemisphere negative nitrate aerosol forcing in one CMIP5 model (see Section 3.3.2) caused an unrealistic hemispheric forcing asymmetry in the mid twentieth century.

The methodologies for calculating the offset magnitude and the strength of aerosol forcing were formulated by considering an idealised twentieth-century scenario. It is accepted that real-world aerosol forcing does not appear as a step change, but the best estimates do show a significant increase in forcing around the 1960s (Lamarque et al., 2010). Likewise, the fact the the post-1960 period saw several explosive volcanic eruptions could influence findings.

In Chapter 3 robust physical explanations are given for why the NHML-SHEXT temperature gradient and NHT land mean precipitation are expected to share a similar relationship with global mean temperature as does global mean precipitation. But there is no complete theory regarding why NHML land mean precipitation should show a response to aerosol forcing. Here, it is shown that changes in NHML land mean precipitation can be used an an indicator for changes in global mean precipitation in CMIP5 models. Wu et al. (2013) applied this principle to Northern Hemisphere land mean precipitation changes as a whole, also showing that observed changes in Northern Hemisphere land mean precipitation closely matched modelled changes. This match between observations and models though is a consequence of the good agreement in the tropical region, which dominates the signal hemispherically.

With a clear disagreement between observed and modelled precipitation in the NHML region and a lack of theory concerning precipitation changes it is difficult to determine whether models are poorly simulating important aspects of precipitation change here. Although other drivers, such as the land surface, will undoubtedly influence precipitation change and variability in the NHML region, the precipitation response to aerosol forcing around the 1960s is unquestionable. Very simply, this is expected as cooling of the surface by aerosols (sulphate and black carbon) should increase tropospheric static stability and suppress convection, as should tropospheric heating from black carbon aerosols.

A limitation of Chapter 4 is the use of just one method to detect a breakpoint. This statistical method was limited in that it could only detect a single breakpoint and

it assumed that the inhomogeneity in either the precipitation or runoff time series that caused this breakpoint was instantaneous, rather than transient. However, initial exploratory analysis of the two time series suggested that a split in the runoff-precipitation relationship was a real feature. But, unlike in Chapters 2 and 3 when the offset was always measured at 1960, because this is when aerosol emissions spiked, no prior assumption about the timing of a breakpoint was made. Therefore, it was tested for at multiple locations. Improvements could be made by looking for change points in the individual time series. This might be particularly enlightening at smaller spatial scales, especially after concluding that the North American continent is likely to be the source of the breakpoint.

A frustrating impediment to this work is the lack of suitably long observed evapotranspiration records. Evapotranspiration is approximately the residual in the precipitation and runoff time series. An observed record for this variable would serve to test the conclusions made in this chapter. This work was also limited somewhat by simple assumptions made about changes in water storage. While it is probably valid to assume that natural changes in storage terms are negligible, this might not hold for changes in water storage due to human activities.

5.3 Ideas for future research

Although Chapter 2 showed that global mean precipitation change in the twentieth century is well understood, even in the absence of observations over the oceans, the topic concerning forcing and feedbacks and their role in determining twentieth-century temperature change in models was not satisfactorily concluded. Between the CMIPs it was suggested that the correlation between forcing and feedback would cease to exist as aerosol forcing became more difficult to tune in a GCM. But it would seem that a greater spread in SW forcing in CMIP5 models is offset by a greater spread in LW forcing to narrow the range of net forcing. An interesting future project would investigate reasons for this. Is it still a con-

sequence of deliberate choices made during the modelling process, or a natural outcome of increased complexity? Answering such a question may require insight into the choices made by individual modelling centres. Or it could take the form of several idealised experiments using a single model. By controlling some components and varying others it might be possible to isolate the reason behind this SW forcing/LW forcing anti-correlation.

A key focus of future work should be on improving existing precipitation observations in the NHML region and elsewhere. Various projects attempt to homogenise temperature records in a systematic and coordinated way. Attempts are even being made to improve spatial and temporal coverage of temperature observations by instigating citizen science projects to digitise old weather reports. No such efforts have been made to improve the quality and quantity of records that are routinely used to construct gridded precipitation datasets. Given the conclusions in Chapter 4 (see Section 4.3.3) a focus on the North American continent is warranted. Further, while the suitability of reanalyses was questioned in Section 1.2.3, there might be value in some datasets, especially as new or improved products become available. Reanalyses bridge the gap between observations and models and it would be interesting to see whether any products contain an obvious precipitation response to aerosol forcing, across a range of regions and scales.

It would be interesting to reconcile the work of Gedney et al. (2014) with the work in Chapter 4. They showed that aerosols not only affect precipitation but have a detectable effect on the observed land water budget. Increased aerosol loading around the 1960s was shown to increase observed NHML river flow. Although Chapter 4 considers the effect of many direct and indirect human influences, arguing that they should not cause large scale breakpoints or introduce non-linearities in the runoff-precipitation relationship, it does not directly address this particular influence. This could perhaps explain the remaining discrepancy that cannot be accounted for, because this effect on runoff may have masked decreases expected because of aerosol-driven decreases in precipitation. Access to the data of Ged-

ney et al. (2014) (their study considers many of the same river catchments as in Chapter 4) could help to remove this effect from observed runoff. Further, a similar offline hydrological model setup could be used to remove, for example, CO_2 physiological effects and even climate effects (mostly temperature) from the runoff time series. Removing these separate effects would require the assumption that they are linearly additive. This is perhaps valid when using offline models, but is less likely to be valid when considering the fully coupled climate system, because of numerous feedbacks between the land surface and atmosphere.

5.4 Contribution to the research field

This thesis has contributed towards the research field in several ways. Firstly, it has furthered understanding of changes in global mean precipitation, especially in the twentieth century, transient climate. Hopefully this can inform research into global mean precipitation projections and improve understanding of the implication of various mitigation (and even geoengineering) scenarios.

However, returning to the adage from Chapter 1 that nobody lives in the global mean, this thesis has improved understanding of regional land mean precipitation change. This has been achieved by building a robust framework linking NHML land precipitation, NHML runoff and NHML aerosol forcing. In Chapter 3, NHML land mean precipitation in models is shown to respond in a predictable way to mid-twentieth-century aerosol forcing. But, noticeably, observed NHML land mean precipitation does not show an aerosol response and is not in keeping with this framework.

The use of observed runoff has been key in strengthening the findings in this thesis. This would seem to the first study to use runoff changes to verify precipitation changes in this way. It is acknowledged that changes in evapotranspiration and even catchment water storage need to be carefully considered when using such an approach. But employing complementary hydrological data plays a key role in

strengthening conclusions.

Overall, it is likely that this inconsistency can be attributed to inhomogeneities in these widely-used observed precipitation datasets, especially in the early twentieth century. This has strong implications for research that uses these observations and it is recommended that caution should be taken prior to drawing conclusions based on them.

Bibliography

J. C. Adam and D. P. Lettenmaier. Adjustment of global gridded precipitation for systematic bias. *J. Geophys. Res.*, 108:4257, 2003.

J. C. Adam and D. P. Lettenmaier. Application of new precipitation and reconstructed streamflow products to streamflow trend attribution in Northern Eurasia. *J. Climate*, 21:1807–1828, 2008. doi: 10.1175/2007JCLI1535.1.

J. C. Adam, E. A. Clark, D. P. Lettenmaier, and E. F. Wood. Correction of global precipitation products for orographic effects. *J. Climate*, 19:15–38, 2006. doi: 10.1175/JCLI3604.1.

J. C. Adam, I. Haddeland, F. Su, and D. P. Lettenmaier. Simulation of reservoir influences on annual and seasonal streamflow changes for the Lena, Yenisei, and Ob' rivers. *J. Geophys. Res.*, 112:D24114, 2007. doi: 10.1029/2007JD008525.

R. F. Adler, G. J. Huffman, A. Chang, R. Ferraro, P.-P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin. The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeor.*, 4:1147–1167, 2003.

A. Ahlström, G. Schurgers, A. Arneth, and B. Smith. Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. *Environ. Res. Lett.*, 7:044008, 2012.

B. A. Albrecht. Aerosols, cloud microphysics, and fractional cloudiness. *Science*, 245:1227–1230, 1989. doi: 10.1126/science.245.4923.1227.

L. V. Alexander and P. D. Jones. Updated precipitation series for the U.K. and discussion of recent extremes. *Atmos. Sci. Lett.*, 1:142–150, 2000. doi: 10.1006/asle.2000.0016.

R. P. Allan, C. Liu, M. Zahn, D. A. Lavers, E. Koukouvagias, and A. Bodas-Salcedo. Physically consistent responses of the global atmospheric hydrological cycle in models and observations. *Surv. Geophys.*, 35:533–552, 2014. doi: 10.1007/s10712-012-9213-z.

M. R. Allen and W. J. Ingram. Constraints on future changes in climate and the hydrologic cycle. *Nature*, 419:224–232, 2002. doi: 10.1038/nature01092.

R. J. Allen, J. R. Norris, and M. Wild. Evaluation of multidecadal variability in CMIP5 surface solar radiation and inferred underestimation of aerosol direct effects over Europe, China, Japan, and India. *J. Geophys. Res.*, 118:6311–6336, 2013. doi: 10.1002/jgrd.50426.

B. T. Anderson, J. R. Knight, M. A. Ringer, C. Deser, A. S. Phillips, J.-H. Yoon, and A. Cherchi. Climate forcings and climate sensitivities diagnosed from atmospheric global circulation models. *Climate Dyn.*, 35:1461–1475, 2010. doi: 10.1007/s00382-010-0798-y.

T. Andrews, P. M. Forster, and J. M. Gregory. A surface energy perspective on climate change. *J. Climate*, 22:2557–2570, 2009.

T. Andrews, P. M. Forster, O. Boucher, N. Bellouin, and A. Jones. Precipitation, radiative forcing and global temperature change. *Geophys. Res. Lett.*, 37: L14701, 2010.

T. Andrews, J. M. Gregory, M. J. Webb, and K. E. Taylor. Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.*, 39:L09712, 2012. doi: 10.1029/2012GL051607.

T. Andrews, J. M. Gregory, and M. J. Webb. The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Climate*, 28:1630–1648, 2015. doi: 10.1175/JCLI-D-14-00545.1.

A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese. A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data*, 5:71–99, 2013.

N. Bellouin, A. Jones, J. Haywood, and S. A. Christopher. Updated estimate of aerosol direct radiative forcing from satellite observations and comparison against the Hadley Centre climate model. *J. Geophys. Res.*, 113:D10205, 2008. doi: 10.1029/2007JD009385.

P. Berg, C. Moseley, and J. O. Haerter. Strong increase in convective precipitation in response to higher temperatures. *Nature Geoscience*, 6:181–185, 2013. doi: 10.1038/ngeo1731.

W. R. Berghuijs, R. A. Woods, and M. Hrachowitz. A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature Climate Change*, 4:583–586, 2014. doi: 10.1038/nclimate2246.

M. J. Best, M. Pryor, D. B. Clark, G. G. Rooney, R. L. H. Essery, C. B. Ménard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding. The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geosci. Model Dev.*, 4:677–699, 2011. doi: 10.5194/gmd-4-677-2011.

H. Biemans, R. W. A. Hutjes, P. Kabat, B. J. Strengers, D. Gerten, and S. Rost. Effects of precipitation uncertainty on discharge calculations for main river basins. *J. Hydrometeor.*, 10:1011–1025, 2009. doi: 10.1175/2008JHM1067.1.

N. L. Bindoff, P. Stott, K. M. AchutaRao, M. R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I. I. Mokhov, J. Overland, J. Perlwitz, R. Sebbari, and X. Zhang. Detection and attribution of climate change: from

global to regional. In T. F. Stocker et al., editors, *Climate Change 2013: The Physical Science Basis*, pages 867–952. Cambridge University Press, 2013.

G. Boer and B. Yu. Climate sensitivity and response. *Climate Dyn.*, 20:415–429, 2003. doi: 10.1007/s00382-002-0283-3.

M. A. Bollasina, Y. Ming, and V. Ramaswamy. Anthropogenic aerosols and the weakening of the South Asian summer monsoon. *Science*, 334:502–505, 2011. doi: 10.1126/science.1204994.

T. C. Bond, S. J. Doherty, D. W. Fahey, P. M. Forster, T. Berntsen, B. J. DeAngelo, M. G. Flanner, S. Ghan, B. Krcher, D. Koch, S. Kinne, Y. Kondo, P. K. Quinn, M. C. Sarofim, M. G. Schultz, M. Schulz, C. Venkataraman, H. Zhang, S. Zhang, N. Bellouin, S. K. Guttikunda, P. K. Hopke, M. Z. Jacobson, J. W. Kaiser, Z. Klimont, U. Lohmann, J. P. Schwarz, D. Shindell, T. Storelvmo, S. G. Warren, and C. S. Zender. Bounding the role of black carbon in the climate system: A scientific assessment. *J. Geophys. Res.*, 118:5380–5552, 2013. doi: 10.1002/jgrd.50171.

B. B. B. Booth, N. J. Dunstone, P. R. Halloran, T. Andrews, and N. Bellouin. Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, 484:228–232, 2012. doi: 10.1038/nature10946.

O. Boucher, D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S. K. Satheesh, S. Sherwood, B. Stevens, and X. Y. Zhang. Clouds and aerosols. In T. F. Stocker et al., editors, *Climate Change 2013: The Physical Science Basis*, pages 571–658. Cambridge University Press, 2013.

M. I. Budyko. *Climate and Life*. Academic Press, New York, 1974.

R. Chadwick, I. Boutle, and G. Martin. Spatial patterns of precipitation change in CMIP5: Why the rich do not get richer in the tropics. *J. Climate*, 26:3803–3822, 2013. doi: 10.1175/JCLI-D-12-00543.1.

C.-Y. Chang, J. C. H. Chiang, M. F. Wehner, A. R. Friedman, and R. Ruedy. Sulfate aerosol control of tropical atlantic climate over the twentieth century. *J. Climate*, 24:2540–2555, 2011. doi: 10.1175/2010JCLI4065.1.

R. Cherian, J. Quaas, M. Salzmann, and M. Wild. Pollution trends over Europe constrain global aerosol forcing as simulated by climate models. *Geophys. Res. Lett.*, 41:2176–2181, 2014. doi: 10.1002/2013GL058715.

C. Chou, J. D. Neelin, C.-A. Chen, and J.-Y. Tu. Evaluating the rich-get-richer mechanism in tropical precipitation change under global warming. *J. Climate*, 22:1982–2005, 2009. doi: 10.1175/2008JCLI2471.1.

D. B. Clark, L. M. Mercado, S. Sitch, C. D. Jones, N. Gedney, M. J. Best, M. Pryor, G. G. Rooney, R. L. H. Essery, E. Blyth, O. Boucher, R. J. Harding, C. Huntingford, and P. M. Cox. The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics. *Geosci. Model Dev.*, 4:701–722, 2011. doi: 10.5194/gmd-4-701-2011.

M. Collins, S.-I. An, W. Cai, A. Ganachaud, E. Guilyardi, F.-F. Jin, M. Jochum, M. Lengaigne, S. Power, A. Timmermann, G. Vecchi, and A. Wittenberg. The impact of global warming on the tropical Pacific Ocean and El Niño. *Nature Geoscience*, 3:391–397, 2010. doi: 10.1038/ngeo868.

G. P. Compo, J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, R. S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, Ø. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff, and S. J. Worley. The twentieth century reanalysis project. *Quart. J. Roy. Meteor. Soc.*, 137:1–28, 2011. doi: 10.1002/qj.776.

K. Cowtan and R. G. Way. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, 140: 1935–1944, 2014. doi: 10.1002/qj.2297.

A. Dai. Precipitation characteristics in eighteen coupled climate models. *J. Climate*, 19:4605–4630, 2006. doi: 10.1175/JCLI3884.1.

A. Dai, T. Qian, K. E. Trenberth, and J. D. Milliman. Changes in continental freshwater discharge from 1948 to 2004. *J. Climate*, 22:2773–2792, 2009. doi: 10.1175/2008JCLI2592.1.

D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thópaut, and F. Vitart. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, 137:553–597, 2011. doi: 10.1002/qj.828.

F. Dentener, S. Kinne, T. Bond, O. Boucher, J. Cofala, S. Generoso, P. Ginoux, S. Gong, J. J. Hoelzemann, A. Ito, L. Marelli, J. E. Penner, J.-P. Putaud, C. Textor, M. Schulz, G. R. van der Werf, and J. Wilson. Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom. *Atmos. Chem. Phys.*, 6:4321–4344, 2006. doi: 10.5194/acp-6-4321-2006.

S. H. Derbyshire, A. V. Maidens, S. F. Milton, R. A. Stratton, and M. R. Willett. Adaptive detrainment in a convective parametrization. *Quart. J. Roy. Meteor. Soc.*, 137:1856–1871, 2011. doi: 10.1002/qj.875.

Y. Ding, D. Yang, B. Ye, and N. Wang. Effects of bias correction on precipitation trend over china. *J. Geophys. Res.*, 112:D13116, 2007.

L. J. Donner, B. L. Wyman, R. S. Hemler, L. W. Horowitz, Y. Ming, M. Zhao, J.-C. Golaz, P. Ginoux, S.-J. Lin, M. D. Schwarzkopf, J. Austin, G. Alaka, W. F. Cooke, T. L. Delworth, S. M. Freidenreich, C. T. Gordon, S. M. Griffies, I. M.

Held, W. J. Hurlin, S. A. Klein, T. R. Knutson, A. R. Langenhorst, H.-C. Lee, Y. Lin, B. I. Magi, S. L. Malyshev, P. C. D. Milly, V. Naik, M. J. Nath, R. Pincus, J. J. Ploshay, V. Ramaswamy, C. J. Seman, E. Shevliakova, J. J. Sirutis, W. F. Stern, R. J. Stouffer, R. J. Wilson, M. Winton, A. T. Wittenberg, and F. Zeng. The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *J. Climate*, 24:3484–3519, 2011. doi: 10.1175/2011JCLI3955.1.

J.-L. Dufresne and S. Bony. An assessment of the primary sources of spread of global warming estimates from coupled atmosphere-ocean models. *J. Climate*, 21:5135–5144, 2008. doi: 10.1175/2008JCLI2239.1.

E. E. Ebert, J. E. Janowiak, and C. Kidd. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, 88:47–64, 2007. doi: 10.1175/BAMS-88-1-47.

N. Elguindi, S. Somot, M. Déqué, and W. Ludwig. Climate change evolution of the hydrological balance of the Mediterranean, Black and Caspian Seas: impact of climate model resolution. *Climate Dyn.*, 36:205–228, 2011. doi: 10.1007/s00382-009-0715-4.

P. Falloon, R. Betts, A. Wiltshire, R. Dankers, C. Mathison, D. McNeall, P. Bates, and M. Trigg. Validation of river flows in HadGEM1 and HadCM3 with the TRIP river flow model. *J. Hydrometeor.*, 12:1157–1180, 2011. doi: 10.1175/2011JHM1388.1.

B. M. Fekete, C. J. Vörösmarty, and W. Grabs. High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochem. Cycles*, 16:1042, 2002. doi: 10.1029/1999GB001254.

C. R. Ferguson and G. Villarini. Detecting inhomogeneities in the twentieth century reanalysis over the central United States. *J. Geophys. Res.*, 117:D05123, 2012. doi: 10.1029/2011JD016988.

A. J. Ferraro, E. J. Highwood, and A. J. Charlton-Perez. Weakened tropical circulation and reduced precipitation in response to geoengineering. *Environ. Res. Lett.*, 9:014001, 2014.

J. B. Fisher, M. Sikka, S. Sitch, P. Ciais, B. Poulter, D. Galbraith, J.-E. Lee, C. Huntingford, N. Viovy, N. Zeng, A. Ahlström, M. R. Lomas, P. E. Levy, C. Frankenberg, S. Saatchi, and Y. Malhi. African tropical rainforest net carbon dioxide fluxes in the twentieth century. *Philos. Trans. Roy. Soc. B*, 368: 20120376, 2013. doi: 10.1098/rstb.2012.0376.

P. Forster, V. Ramaswamy, P. Artaxo, T. Berntsen, R. Betts, D. W. Fahey, J. Haywood, J. Lean, D. C. Lowe, G. Myhre, J. Nganga, R. Prinn, G. Raga, M. Schulz, and R. Van Dorland. Changes in atmospheric constituents and in radiative forcing. In S. Solomon et al., editors, *Climate Change 2007: The Physical Science Basis*, pages 129–234. Cambridge University Press, 2007.

P. M. Forster and K. E. Taylor. Climate forcings and climate sensitivities diagnosed from coupled climate model integrations. *J. Climate*, 19:6181–6194, 2006. doi: 10.1175/JCLI3974.1.

P. M. Forster, T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka. Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res.*, 118:1139–1150, 2013.

H. J. Fowler and C. G. Kilsby. Implications of changes in seasonal and annual extreme rainfall. *Geophys. Res. Lett.*, 30:1720, 2003. doi: 10.1029/2003GL017327.

A. R. Friedman, Y.-T. Hwang, J. C. H. Chiang, and D. M. W. Frierson. Interhemispheric temperature asymmetry over the twentieth century and in future projections. *J. Climate*, 26:5419–5433, 2013. doi: 10.1175/JCLI-D-12-00525.1.

K. Frieler, M. Meinshausen, T. Schneider von Deimling, T. Andrews, and P. Forster. Changes in global-mean precipitation in response to warming, greenhouse gas forcing and black carbon. *Geophys. Res. Lett.*, 38:L04702, 2011.

B. Fu. On the calculation of the evaporation from land surface (in Chinese). *Sci. Atmos. Sin.*, 1:23–31, 1981.

J. C. Fyfe, N. P. Gillett, and F. W. Zwiers. Overestimated global warming over the past 20 years. *Nature Climate Change*, 3:767–769, 2013. doi: doi:10.1038/nclimate1972.

N. Gedney, P. M. Cox, R. A. Betts, O. Boucher, C. Huntingford, and P. A. Stott. Detection of a direct carbon dioxide effect in continental river runoff records. *Nature*, 439:835–838, 2006. doi: 10.1038/nature04504.

N. Gedney, C. Huntingford, G. P. Weedon, N. Bellouin, O. Boucher, and P. M. Cox. Detection of solar dimming and brightening effects on Northern Hemisphere river flow. *Nature Geoscience*, 7:796–800, 2014. doi: 10.1038/ngeo2263.

D. Gerten, S. Rost, W. von Bloh, and W. Lucht. Causes of change in 20th century global river discharge. *Geophys. Res. Lett.*, 35:L20405, 2008. doi: 10.1029/2008GL035258.

S. J. Ghan, H. Abdul-Razzak, A. Nenes, Y. Ming, X. Liu, M. Ovchinnikov, B. Shipway, N. Meskhidze, J. Xu, and X. Shi. Droplet nucleation: Physically-based parameterizations and comparative evaluation. *J. Adv. Model. Earth Syst.*, 3:M10001, 2011. doi: 10.1029/2011MS000074.

N. P. Gillett, A. J. Weaver, F. W. Zwiers, and M. F. Wehner. Detection of volcanic influence on global precipitation. *Geophys. Res. Lett.*, 31:L12217, 2004.

P. Good, W. Ingram, F. H. Lambert, J. A. Lowe, J. M. Gregory, M. J. Webb, M. A. Ringer, and P. Wu. A step-response approach for predicting and understanding non-linear precipitation changes. *Climate Dyn.*, 39:2789–2803, 2012. doi: 10.1007/s00382-012-1571-1.

L. J. Gordon, W. Steffen, B. F. Jönsson, C. Folke, M. Falkenmark, and Å. Johannessen. Human modification of global water vapor flows from the land surface. *Proc. Natl. Acad. Sci. (USA)*, 102:7612–7617, 2005. doi: 10.1073/pnas.0500208102.

D. Gregory and P. R. Rowntree. A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Wea. Rev.*, 118:1483–1506, 1990.

J. Gregory and M. Webb. Tropospheric adjustment induces a cloud component in CO_2 forcing. *J. Climate*, 21:58–71, 2008. doi: 10.1175/2007JCLI1834.1.

J. M. Gregory and P. M. Forster. Transient climate response estimated from radiative forcing and observed temperature change. *J. Geophys. Res.*, 113: D23105, 2008. doi: 10.1029/2008JD010405.

J. M. Gregory, W. J. Ingram, M. A. Palmer, G. S. Jones, P. A. Stott, R. B. Thorpe, J. A. Lowe, T. C. Johns, and K. D. Williams. A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, 31:L03205, 2004.

P. Greve, B. Orlowsky, B. Mueller, J. Sheffield, M. Reichstein, and S. I. Seneviratne. Global assessment of trends in wetting and drying over land. *Nature Geoscience*, 7:716–721, 2014.

P. Greve, L. Gudmundsson, B. Orlowsky, and S. Seneviratne. Introducing a probabilistic Budyko framework. *Geophys. Res. Lett.*, pages n/a–n/a, 2015. doi: 10.1002/2015GL063449.

P. Y. Groisman and E. Y. Rankova. Precipitation trends over the Russian permafrost-free zone: Removing the artifacts of pre-processing. *Int. J. Climatol.*, 21:657–678, 2001.

P. Y. Groisman, V. V. Koknaeva, T. A. Belokrylova, and T. R. Karl. Overcoming biases of precipitation measurement: A history of the USSR experience. *Bull. Amer. Meteor. Soc.*, 72:1725–1733, 1991.

S. Hagemann and L. Dümenil. A parametrization of the lateral waterflow for the global scale. *Climate Dyn.*, 14:17–31, 1998. doi: 10.1007/s003820050205.

J. Hannaford and T. J. Marsh. High-flow and flood trends in a network of undis-

turbed catchments in the UK. *Int. J. Climatol.*, 28:1325–1338, 2008. doi: 10.1002/joc.1643.

J. Hansen, M. Sato, R. Ruedy, L. Nazarenko, A. Lacis, G. A. Schmidt, G. Russell, I. Aleinov, M. Bauer, S. Bauer, N. Bell, B. Cairns, V. Canuto, M. Chandler, Y. Cheng, A. Del Genio, G. Faluvegi, E. Fleming, A. Friend, T. Hall, C. Jackman, M. Kelley, N. Kiang, D. Koch, J. Lean, J. Lerner, K. Lo, S. Menon, R. Miller, P. Minnis, T. Novakov, V. Oinas, J. Perlwitz, J. Perlwitz, D. Rind, A. Romanou, D. Shindell, P. Stone, S. Sun, N. Tausnev, D. Thresher, B. Wielicki, T. Wong, M. Yao, and S. Zhang. Efficacy of climate forcings. *J. Geophys. Res.*, 110:D18104, 2005. doi: 10.1029/2005JD005776.

R. J. Harding, G. P. Weedon, H. A. van Lanen, and D. B. Clark. The future for global water assessment. *J. Hydrol.*, 518:186–193, 2014. doi: 10.1016/j.jhydrol.2014.05.014.

I. Harris, P. Jones, T. Osborn, and D. Lister. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *Int. J. Climatol.*, 34: 623–642, 2014. doi: 10.1002/joc.3711.

J. Haywood and O. Boucher. Estimates of the direct and indirect radiative forcing due to tropospheric aerosols: A review. *Rev. Geophys.*, 38:513–543, 2000.

J. M. Haywood, A. Jones, N. Bellouin, and D. Stephenson. Asymmetric forcing from stratospheric aerosols impacts Sahelian rainfall. *Nature Climate Change*, 3:660–665, 2013. doi: 10.1038/nclimate1857.

G. C. Hegerl, E. Black, R. P. Allan, W. J. Ingram, D. Polson, K. E. Trenberth, R. S. Chadwick, P. A. Arkin, B. B. Sarojini, A. Becker, A. Dai, P. J. Durack, D. Easterling, H. J. Fowler, E. J. Kendon, G. J. Huffman, C. Liu, R. Marsh, M. New, T. J. Osborn, N. Skliris, P. A. Stott, P.-L. Vidale, S. E. Wijffels, L. J. Wilcox, K. M. Willett, and X. Zhang. Challenges in quantifying changes in the global water cy cle. *Bull. Amer. Meteor. Soc.*, 2014. doi: 10.1175/BAMS-D-13-00212.1.

I. M. Held and B. J. Soden. Robust responses of the hydrological cycle to global warming. *J. Climate*, 19:5686–5699, 2006. doi: 10.1175/JCLI3990.1.

Y. Hirabayashi, R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae. Global flood risk under climate change. *Nature Climate Change*, 3:816–821, 2013. doi: 10.1038/nclimate1911.

N. Hsu, S.-C. Tsay, M. King, and J. Herman. Deep blue retrievals of Asian aerosol properties during ACE-Asia. *IEEE Trans. Geosci. Remote Sens.*, 44:3180–3195, 2006. doi: 10.1109/TGRS.2006.879540.

G. J. Huffman, D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong, K. P. Bowman, and E. F. Stocker. The TRMM multisatellite precipitation analysis (TMPA): Quasi-Global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, 8:38–55, 2007. doi: 10.1175/JHM560.1.

C. Huntingford, P. M. Cox, L. M. Mercado, S. Sitch, N. Bellouin, O. Boucher, and N. Gedney. Highly contrasting effects of different climate forcing agents on terrestrial ecosystem services. *Philos. Trans. Roy. Soc. A*, 369:2026–2037, 2011. doi: 10.1098/rsta.2010.0314.

C. Huntingford, T. Marsh, A. A. Scaife, E. J. Kendon, J. Hannaford, A. L. Kay, M. Lockwood, C. Prudhomme, N. S. Reynard, S. Parry, J. A. Lowe, J. A. Screen, H. C. Ward, M. Roberts, P. A. Stott, V. A. Bell, M. Bailey, A. Jenkins, T. Legg, F. E. L. Otto, N. Massey, N. Schaller, J. Slingo, and M. R. Allen. Potential influences on the United Kingdom's floods of winter 2013/14. *Nature Climate Change*, 4:769–777, 2014.

T. G. Huntington. CO_2-induced suppression of transpiration cannot explain increasing runoff. *Hydrol. Processes*, 22:311–314, 2008. doi: 10.1002/hyp.6925.

J. W. Hurrell. Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, 269:676–679, 1995. doi: 10.1126/science.269.5224.676.

Y.-T. Hwang, D. M. W. Frierson, and S. M. Kang. Anthropogenic sulfate aerosol and the southward shift of tropical precipitation in the late 20th century. *Geophys. Res. Lett.*, 40:2845–2850, 2013.

C. E. Iles, G. C. Hegerl, A. P. Schurer, and X. Zhang. The effect of volcanic eruptions on global precipitation. *J. Geophys. Res.*, 118:8770–8786, 2013. doi: 10.1002/jgrd.50678.

M. R. Jones, H. J. Fowler, C. G. Kilsby, and S. Blenkinsop. An assessment of changes in seasonal and annual extreme rainfall in the UK between 1961 and 2009. *Int. J. Climatol.*, 33:1178–1194, 2013. doi: 10.1002/joc.3503.

I. W. Jung, H. Chang, and J. Risley. Effects of runoff sensitivity and catchment characteristics on regional actual evapotranspiration trends in the conterminous US. *Environ. Res. Lett.*, 8:044002, 2013. doi: 10.1088/1748-9326/8/4/044002.

M. Jung, M. Reichstein, P. Ciais, S. I. Seneviratne, J. Sheffield, M. L. Goulden, G. Bonan, A. Cescatti, J. Chen, R. de Jeu, A. J. Dolman, W. Eugster, D. Gerten, D. Gianelle, N. Gobron, J. Heinke, J. Kimball, B. E. Law, L. Montagnani, Q. Mu, B. Mueller, K. Oleson, D. Papale, A. D. Richardson, O. Roupsard, S. Running, E. Tomelleri, N. Viovy, U. Weber, C. Williams, E. Wood, S. Zaehle, and K. Zhang. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, 467:951–954, 2010. doi: 10.1038/nature09396.

E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph. The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, 77:437–471, 1996.

T. R. Karl, R. G. Quayle, and P. Y. Groisman. Detecting climate variations and change: New challenges for observing and data management systems. *J. Climate*, 6:1481–1494, 1993.

J. T. Kiehl. Twentieth century climate model response and climate sensitivity. *Geophys. Res. Lett.*, 34:L22710, 2007. doi: 10.1029/2007GL031383.

K. Klein Goldewijk and P. H. Verburg. Uncertainties in global-scale reconstructions of historical land use: an illustration using the HYDE data set. *Landscape Ecol.*, 28:861–877, 2013. doi: 10.1007/s10980-013-9877-x.

N. P. Klingaman and S. J. Woolnough. The role of air-sea coupling in the simulation of the Madden-Julian oscillation in the Hadley Centre model. *Quart. J. Roy. Meteor. Soc.*, 140:2272–2286, 2014. doi: 10.1002/qj.2295.

R. Knutti. Why are climate models reproducing the observed global surface warming so well? *Geophys. Res. Lett.*, 35:L18704, 2008.

R. Knutti and J. Sedlacek. Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, 3:369–373, 2013. doi: 10.1038/nclimate1716.

R. Knutti, D. Masson, and A. Gettelman. Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.*, 40:1194–1199, 2013. doi: 10.1002/grl.50256.

Y. Kosaka and S.-P. Xie. Recent global-warming hiatus tied to equatorial pacific surface cooling. *Nature*, 501:403–407, 2013. doi: doi:10.1038/nature12534.

R. D. Koster, P. A. Dirmeyer, Z. Guo, G. Bonan, E. Chan, P. Cox, C. T. Gordon, S. Kanae, E. Kowalczyk, D. Lawrence, P. Liu, C. Lu, S. Malyshev, B. McAvaney, K. Mitchell, D. Mocko, T. Oki, K. Oleson, A. Pitman, Y. C. Sud, C. M. Taylor, D. Verseghy, R. Vasic, Y. Xue, and T. Yamada. Regions of strong coupling between soil moisture and precipitation. *Science*, 305:1138–1140, 2004.

N. Y. Krakauer and I. Fung. Mapping and attribution of change in streamflow in the coterminous United States. *Hydrol. Earth Syst. Sci.*, 12:1111–1120, 2008. doi: 10.5194/hess-12-1111-2008.

T. Kühn, A.-I. Partanen, A. Laakso, Z. Lu, T. Bergman, S. Mikkonen, H. Kokkola, H. Korhonen, P. Risnen, D. G. Streets, S. Romakkaniemi, and A. Laaksonen. Climate impacts of changing aerosol emissions since 1996. *Geophys. Res. Lett.*, 41:4711–4718, 2014. doi: 10.1002/2014GL060349.

H. L. Kuo. On formation and intensification of tropical cyclones through latent heat release by cumulus convection. *J. Atmos. Sci.*, 22:40–63, 1965.

M. M. Kvalevåg, B. H. Samset, and G. Myhre. Hydrological sensitivity to greenhouse gases and aerosols in a global climate model. *Geophys. Res. Lett.*, 40: 1432–1438, 2013. doi: 10.1002/grl.50318.

D. Labat, Y. Goddéris, J. L. Probst, and J. L. Guyot. Evidence for global runoff increase related to climate warming. *Adv. Water Resour.*, 27:631–642, 2004. doi: 10.1016/j.advwatres.2004.02.020.

J.-F. Lamarque, T. C. Bond, V. Eyring, C. Granier, A. Heil, Z. Klimont, D. Lee, C. Liousse, A. Mieville, B. Owen, M. G. Schultz, D. Shindell, S. J. Smith, E. Stehfest, J. Van Aardenne, O. R. Cooper, M. Kainuma, N. Mahowald, J. R. McConnell, V. Naik, K. Riahi, and D. P. van Vuuren. Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmos. Chem. Phys.*, 10:7017–7039, 2010.

F. H. Lambert and M. R. Allen. Are changes in global precipitation constrained by the tropospheric energy budget? *J. Climate*, 22:499–517, 2009. doi: 10.1175/2008JCLI2135.1.

F. H. Lambert and N. E. Faull. Tropospheric adjustment: The response of two general circulation models to a change in insolation. *Geophys. Res. Lett.*, 34: L03701, 2007.

F. H. Lambert and M. J. Webb. Dependency of global mean precipitation on surface temperature. *Geophys. Res. Lett.*, 35:L16706, 2008.

F. H. Lambert, P. A. Stott, M. R. Allen, and M. A. Palmer. Detection and attribution of changes in 20th century land precipitation. *Geophys. Res. Lett.*, 31:L10203, 2004.

F. H. Lambert, A. R. Stine, N. Y. Krakauer, and J. C. H. Chiang. How much will precipitation increase with global warming? *Eos, Trans. Amer. Geophys. Union*, 89:193–194, 2008.

F. H. Lambert, M. J. Webb, M. Yoshimori, and T. Yokohata. The cloud radiative effect on the atmospheric energy budget and global mean precipitation. *Climate Dyn.*, pages 1–25, 2014. doi: 10.1007/s00382-014-2174-9.

C. Le Quéré, G. P. Peters, R. J. Andres, R. M. Andrew, T. A. Boden, P. Ciais, P. Friedlingstein, R. A. Houghton, G. Marland, R. Moriarty, S. Sitch, P. Tans, A. Arneth, A. Arvanitis, D. C. E. Bakker, L. Bopp, J. G. Canadell, L. P. Chini, S. C. Doney, A. Harper, I. Harris, J. I. House, A. K. Jain, S. D. Jones, E. Kato, R. F. Keeling, K. Klein Goldewijk, A. Körtzinger, C. Koven, N. Lefèvre, F. Maignan, A. Omar, T. Ono, G.-H. Park, B. Pfeil, B. Poulter, M. R. Raupach, P. Regnier, C. Rödenbeck, S. Saito, J. Schwinger, J. Segschneider, B. D. Stocker, T. Takahashi, B. Tilbrook, S. van Heuven, N. Viovy, R. Wanninkhof, A. Wiltshire, and S. Zaehle. Global carbon budget 2013. *Earth Syst. Sci. Data*, 6:235–263, 2014. doi: 10.5194/essd-6-235-2014.

D. R. Legates. Global and terrestrial precipitation: A comparative assessment of existing climatologies. *Int. J. Climatol.*, 15:237–258, 1995.

E. M. Leibensperger, L. J. Mickley, D. J. Jacob, W.-T. Chen, J. H. Seinfeld, A. Nenes, P. J. Adams, D. G. Streets, N. Kumar, and D. Rind. Climatic effects of 1950-2050 changes in US anthropogenic aerosols – Part 2: Climate response. *Atmos. Chem. Phys.*, 12:3349–3362, 2012. doi: 10.5194/acp-12-3349-2012.

G. Lenderink and E. van Meijgaard. Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geoscience*, 1:511–514, 2008. doi: 10.1038/ngeo262.

H. Lin, G. Brunet, and J. S. Fontecilla. Impact of the Madden-Julian Oscillation on the intraseasonal forecast skill of the North Atlantic Oscillation. *Geophys. Res. Lett.*, 37:L19803, 2010. doi: 10.1029/2010GL044315.

R. Lindsay, M. Wensnahan, A. Schweiger, and J. Zhang. Evaluation of seven different atmospheric reanalysis products in the arctic. *J. Climate*, 27:2588–2606, 2014. doi: 10.1175/JCLI-D-13-00014.1.

U. Lohmann and E. Roeckner. Design and performance of a new cloud microphysics scheme developed for the ECHAM general circulation model. *Climate Dyn.*, 12:557–572, 1996. doi: 10.1007/BF00207939.

C. Lorenz and H. Kunstmann. The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *J. Hydrometeor.*, 13:1397–1420, 2012. doi: 10.1175/JHM-D-11-088.1.

F. F. Malavelle, J. M. Haywood, P. R. Field, A. A. Hill, S. J. Abel, A. P. Lock, B. J. Shipway, and K. McBeath. A method to represent subgrid-scale updraft velocity in kilometer-scale models: Implication for aerosol activation. *J. Geophys. Res.*, 119:4149–4173, 2014. doi: 10.1002/2013JD021218.

S. Manabe and R. T. Wetherald. The effects of doubling the CO_2 concentration on the climate of a General Circulation Model. *J. Atmos. Sci.*, 32:3–15, 1975.

B. Marzeion, J. G. Cogley, K. Richter, and D. Parkes. Attribution of global glacier mass loss to anthropogenic and natural causes. *Science*, 345:919–921, 2014. doi: 10.1126/science.1254702.

G. J. McCabe and D. M. Wolock. A step increase in streamflow in the conterminous United States. *Geophys. Res. Lett.*, 29:2185, 2002. doi: 10.1029/2002GL015999.

J. W. McClelland, R. M. Holmes, B. J. Peterson, and M. Stieglitz. Increasing river discharge in the Eurasian Arctic: Consideration of dams, permafrost thaw, and fires as potential agents of change. *J. Geophys. Res.*, 109:D18102, 2004. doi: 10.1029/2004JD004583.

A. Mehran, A. AghaKouchak, and T. J. Phillips. Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. *J. Geophys. Res.*, 119:1695–1707, 2014. doi: 10.1002/2013JD021152.

E. Mekis and L. A. Vincent. An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmos.–Ocean*, 49:163–177, 2011. doi: 10.1080/07055900.2011.583910.

S. Menon, J. Hansen, L. Nazarenko, and Y. Luo. Climate effects of black carbon aerosols in China and India. *Science*, 297:2250–2253, 2002. doi: 10.1126/science.1075159.

R. L. Miller, G. A. Schmidt, L. S. Nazarenko, N. Tausnev, S. E. Bauer, A. D. DelGenio, M. Kelley, K. K. Lo, R. Ruedy, D. T. Shindell, I. Aleinov, M. Bauer, R. Bleck, V. Canuto, Y. Chen, Y. Cheng, T. L. Clune, G. Faluvegi, J. E. Hansen, R. J. Healy, N. Y. Kiang, D. Koch, A. A. Lacis, A. N. LeGrande, J. Lerner, S. Menon, V. Oinas, C. Pérez Garca-Pando, J. P. Perlwitz, M. J. Puma, D. Rind, A. Romanou, G. L. Russell, M. Sato, S. Sun, K. Tsigaridis, N. Unger, A. Voulgarakis, M.-S. Yao, and J. Zhang. CMIP5 historical simulations (1850-2012) with GISS ModelE2. *J. Adv. Model. Earth Syst.*, 6:441–477, 2014. doi: 10.1002/2013MS000266.

J. Milliman, K. Farnsworth, P. Jones, K. Xu, and L. Smith. Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951-2000. *Global Planet. Change*, 62:187–194, 2008. doi: 10.1016/j.gloplacha.2008.03.001.

P. C. D. Milly and K. A. Dunne. Trends in evaporation and surface cooling in the Mississippi River Basin. *Geophys. Res. Lett.*, 28:1219–1222, 2001. doi: 10.1029/2000GL012321.

P. C. D. Milly, K. A. Dunne, and A. V. Vecchia. Global pattern of trends in streamflow and water availability in a changing climate. *Nature*, 438:347–350, 2005. doi: 10.1038/nature04312.

S.-K. Min, X. Zhang, F. W. Zwiers, and G. C. Hegerl. Human contribution to more-intense precipitation extremes. *Nature*, 470:378–381, 2011. doi: 10.1038/nature09763.

Y. Ming, V. Ramaswamy, and G. Persad. Two opposing effects of absorbing aerosols on global-mean precipitation. *Geophys. Res. Lett.*, 37:L13701, 2010.

J. F. B. Mitchell, C. A. Wilson, and W. M. Cunnington. On CO_2 climate sensitivity and model dependence of results. *Quart. J. Roy. Meteor. Soc.*, 113:293–322, 1987.

C. P. Morice, J. J. Kennedy, N. A. Rayner, and P. D. Jones. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set. *J. Geophys. Res.*, 117:D08101, 2012.

G. Myhre. Consistency between satellite-derived and modeled estimates of the direct aerosol effect. *Science*, 325:187–190, 2009. doi: 10.1126/science.1174461.

G. Myhre, D. Shindell, F.-M. Bréon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, T. Takemura, and H. Zhang. Anthropogenic and natural radiative forcing. In T. F. Stocker et al., editors, *Climate Change 2013: The Physical Science Basis*, pages 659–740. Cambridge University Press, 2013.

V. Nešpor and B. Sevruk. Estimation of wind-induced error of rainfall gauge measurements using a numerical simulation. *J. Atmos. Oceanic Technol.*, 16: 450–464, 1999.

M. New, M. Hulme, and P. Jones. Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate. *J. Climate*, 13:2217–2238, 2000.

M. New, M. Todd, M. Hulme, and P. Jones. Precipitation measurements and trends in the twentieth century. *Int. J. Climatol.*, 21:1889–1922, 2001. doi: 10.1002/joc.680.

C. Nilsson, C. A. Reidy, M. Dynesius, and C. Revenga. Fragmentation and flow regulation of the world's large river systems. *Science*, 308:405–408, 2005. doi: 10.1126/science.1107887.

P. O'Gorman, R. Allan, M. Byrne, and M. Previdi. Energetic constraints on precipitation under climate change. *Surv. Geophys.*, 33:585–608, 2012. doi: 10.1007/s10712-011-9159-6.

P. A. O'Gorman. Sensitivity of tropical precipitation extremes to climate change. *Nature Geoscience*, 5:697–700, 2012. doi: 10.1038/ngeo1568.

P. A. O'Gorman and T. Schneider. The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proc. Natl. Acad. Sci. (USA)*, 106:14773–14777, 2009. doi: 10.1073/pnas.0907610106.

K. W. Oleson, D. Lawrence, G. Bonan, B. Drewniak, M. Huang, C. Koven, S. Levis, F. Li, W. Riley, Z. Subin, S. Swenson, P. Thornton, A. Bozbiyik, R. Fisher, C. Heald, E. Kluzek, J.-F. Lamarque, P. Lawrence, L. Leung, W. Lipscomb, S. Muszala, D. Ricciuto, W. Sacks, Y. Sun, J. Tang, and Z.-L. Yang. Technical description of version 4.5 of the Community Land Model (CLM). NCAR Tech. Note NCAR/TN-503+STR, 2013.

P. J. C. Oliveira, E. L. Davin, S. Levis, and S. I. Seneviratne. Vegetation-mediated impacts of trends in global radiation on land hydrology: a global sensitivity study. *Global Change Biol.*, 17:3453–3467, 2011. doi: 10.1111/j.1365-2486.2011.02506.x.

L. Oman, A. Robock, G. L. Stenchikov, and T. Thordarson. High-latitude eruptions cast shadow over the African monsoon and the flow of the Nile. *Geophys. Res. Lett.*, 33:L18711, 2006. doi: 10.1029/2006GL027665.

J. M. Osborne and F. H. Lambert. The missing aerosol response in twentieth-century mid-latitude precipitation observations. *Nature Climate Change*, 4: 374–378, 2014. doi: 10.1038/nclimate2173.

J. M. Osborne, F. H. Lambert, M. Groenendijk, A. B. Harper, C. D. Koven, B. Poulter, T. A. M. Pugh, S. Sitch, B. D. Stocker, A. Wiltshire, and S. Zaehle. Reconciling precipitation with runoff: observed hydrological change in the mid-latitudes. *J. Hydrometeor.*, 2015. doi: 10.1175/JHM-D-15-0055.1.

T. C. Peterson, D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Frland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker. Homogeneity adjustments of in situ atmospheric climate data: a review. *Int. J. Climatol.*, 18:1493–1517, 1998.

S. Piao, P. Friedlingstein, P. Ciais, N. de Noblet-Ducoudré, D. Labat, and S. Zaehle. Changes in climate and land use have a larger direct impact than rising CO_2 on global river runoff trends. *Proc. Natl. Acad. Sci. (USA)*, 104:15242–15247, 2007. doi: 10.1073/pnas.0707213104.

R. S. Plant. A review of the theoretical basis for bulk mass flux convective parameterization. *Atmos. Chem. Phys.*, 10:3529–3544, 2010. doi: 10.5194/acp-10-3529-2010.

D. Polson, G. C. Hegerl, X. Zhang, and T. J. Osborn. Causes of robust seasonal land precipitation changes. *J. Climate*, 26:6679–6697, 2013. doi: 10.1175/JCLI-D-12-00474.1.

M. Previdi. Radiative feedbacks on global precipitation. *Environ. Res. Lett.*, 5: 025211, 2010.

V. Ramanathan and G. Carmichael. Global and regional climate changes due to black carbon. *Nature Geoscience*, 1:221–227, 2008. doi: 10.1038/ngeo156.

V. Ramanathan, C. Chung, D. Kim, T. Bettge, L. Buja, J. T. Kiehl, W. M. Washington, Q. Fu, D. R. Sikka, and M. Wild. Atmospheric brown clouds: Impacts on South Asian climate and hydrological cycle. *Proc. Natl. Acad. Sci. (USA)*, 102:5326–5333, 2005. doi: 10.1073/pnas.0500656102.

J. Reeves, J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu. A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteor. Climatol.*, 46:900–915, 2007. doi: 10.1175/JAM2493.1.

J. C. Rodda and H. Dixon. Rainfall measurement revisited. *Weather*, 67:131–136, 2012. doi: 10.1002/wea.875.

G. H. Roe. Orographic precipitation. *Annu. Rev. Earth Planet. Sci.*, 33:645–671, 2005. doi: 10.1146/annurev.earth.33.092203.122541.

D. Rosenfeld, S. Sherwood, R. Wood, and L. Donner. Climate effects of aerosol-cloud interactions. *Science*, 343:379–380, 2014. doi: 10.1126/science.1247490.

M. Sato, J. E. Hansen, M. P. McCormick, and J. B. Pollack. Stratospheric aerosol optical depths, 1850-1990. *J. Geophys. Res.*, 98:22987–22994, 1993. doi: 10.1029/93JD02553.

J. Scheff and D. M. W. Frierson. Robust future precipitation declines in CMIP5 largely reflect the poleward expansion of model subtropical dry zones. *Geophys. Res. Lett.*, 39:L18704, 2012. doi: 10.1029/2012GL052910.

J. Schewe, J. Heinke, D. Gerten, I. Haddeland, N. W. Arnell, D. B. Clark, R. Dankers, S. Eisner, B. M. Fekete, F. J. Colón-Gonzlez, S. N. Gosling, H. Kim, X. Liu, Y. Masaki, F. T. Portmann, Y. Satoh, T. Stacke, Q. Tang, Y. Wada, D. Wisser, T. Albrecht, K. Frieler, F. Piontek, L. Warszawski, and P. Kabat. Multimodel assessment of water scarcity under climate change. *Proc. Natl. Acad. Sci. (USA)*, 111:3245–3250, 2014. doi: 10.1073/pnas.1222460110.

S. D. Schubert, M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister. On the cause of the 1930s dust bowl. *Science*, 303:1855–1859, 2004. doi: 10.1126/science.1095048.

C. R. Schwalm, D. N. Huntinzger, R. B. Cook, Y. Wei, I. T. Baker, R. P. Neilson, B. Poulter, P. Caldwell, G. Sun, H. Q. Tian, and N. Zeng. A model-data intercomparison of simulated runoff in the contiguous United States: results

from the North America Carbon Regional and Continental Interim-Synthesis. *Biogeosci. Disc.*, 11:1801–1826, 2014. doi: 10.5194/bgd-11-1801-2014.

J. A. Screen and I. Simmonds. The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, 464:1334–1337, 2010. doi: 10.1038/nature09051.

J. A. Screen, I. Simmonds, C. Deser, and R. Tomas. The atmospheric response to three decades of observed Arctic sea ice loss. *J. Climate*, 26:1230–1248, 2013. doi: 10.1175/JCLI-D-12-00063.1.

R. Seager, N. Naik, and G. A. Vecchi. Thermodynamic and dynamic mechanisms for large-scale changes in the hydrological cycle in response to global warming. *J. Climate*, 23:4651–4668, 2010. doi: 10.1175/2010JCLI3655.1.

B. Sevruk. Methods of correction for systematic error in point precipitation measurement for operational use. WMO Oper. Hydrol. Rep. WMO-No 589, World Meteorological Organization: Geneva, 1982.

B. Sevruk and S. Klemm. Catalogue of standard precipitation gauges, instruments and observing methods. WMO Oper. Hydrol. Rep. WMO/TD No. 328, World Meteorological Organization: Geneva, 1989.

T. G. Shepherd. Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, 7:703–708, 2014. doi: doi:10.1038/ngeo2253.

D. Shindell and G. Faluvegi. Climate response to regional radiative forcing during the twentieth century. *Nature Geoscience*, 2:294–300, 2009.

D. T. Shindell. Inhomogeneous forcing and transient climate sensitivity. *Nature Climate Change*, 4:274–277, 2014. doi: 10.1038/nclimate2136.

D. T. Shindell, A. Voulgarakis, G. Faluvegi, and G. Milly. Precipitation response to regional radiative forcing. *Atmos. Chem. Phys.*, 12:6969–6982, 2012.

D. T. Shindell, J.-F. Lamarque, M. Schulz, M. Flanner, C. Jiao, M. Chin, P. J. Young, Y. H. Lee, L. Rotstayn, N. Mahowald, G. Milly, G. Faluvegi, Y. Balkanski, W. J. Collins, A. J. Conley, S. Dalsoren, R. Easter, S. Ghan, L. Horowitz, X. Liu, G. Myhre, T. Nagashima, V. Naik, S. T. Rumbold, R. Skeie, K. Sudo, S. Szopa, T. Takemura, A. Voulgarakis, J.-H. Yoon, and F. Lo. Radiative forcing in the ACCMIP historical and future climate simulations. *Atmos. Chem. Phys.*, 13:2939–2974, 2013. doi: 10.5194/acp-13-2939-2013.

S. Siebert, P. Döll, J. Hoogeveen, J.-M. Faures, K. Frenken, and S. Feick. Development and validation of the global map of irrigation areas. *Hydrol. Earth Syst. Sci.*, 9:535–547, 2005. doi: 10.5194/hess-9-535-2005.

J. Sillmann, V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh. Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res.*, 118:1716–1733, 2013. doi: 10.1002/jgrd.50203.

S. Sitch, B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. O. Kaplan, S. Levis, W. Lucht, M. T. Sykes, K. Thonicke, and S. Venevsky. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biol.*, 9:161–185, 2003. doi: 10.1046/j.1365-2486.2003.00569.x.

S. Sitch, P. Friedlingstein, N. Gruber, S. D. Jones, G. Murray-Tortarolo, A. Ahlström, S. C. Doney, H. Graven, C. Heinze, C. Huntingford, S. Levis, P. E. Levy, M. Lomas, B. Poulter, N. Viovy, S. Zaehle, N. Zeng, A. Arneth, G. Bonan, L. Bopp, J. G. Canadell, F. Chevallier, P. Ciais, R. Ellis, M. Gloor, P. Peylin, S. Piao, C. Le Quéré, B. Smith, Z. Zhu, and R. Myneni. Trends and drivers of regional sources and sinks of carbon dioxide over the past two decades. *Biogeosci. Disc.*, 10:20113–20177, 2013. doi: 10.5194/bgd-10-20113-2013.

B. Smith, I. C. Prentice, and M. T. Sykes. Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting

approaches within European climate space. *Global Ecol. Biogeogr.*, 10:621–637, 2001. doi: 10.1046/j.1466-822X.2001.t01-1-00256.x.

R. N. B. Smith. A scheme for predicting layer clouds and their water content in a general circulation model. *Quart. J. Roy. Meteor. Soc.*, 116:435–460, 1990. doi: 10.1002/qj.49711649210.

S. J. Smith, J. van Aardenne, Z. Klimont, R. J. Andres, A. Volke, and S. Delgado Arias. Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmos. Chem. Phys.*, 11:1101–1116, 2011.

G. L. Stephens, T. L'Ecuyer, R. Forbes, A. Gettlemen, J.-C. Golaz, A. Bodas-Salcedo, K. Suzuki, P. Gabriel, and J. Haynes. Dreary state of precipitation in global models. *J. Geophys. Res.*, 115:D24211, 2010. doi: 10.1029/2010JD014532.

S. M. Sterling, A. Ducharne, and J. Polcher. The impact of global land-cover change on the terrestrial water cycle. *Nature Climate Change*, 3:385–390, 2013. doi: 10.1038/nclimate1690.

D. I. Stern. Reversal of the trend in global anthropogenic sulfur emissions. *Global Environ. Change*, 16:207–220, 2006. doi: http://dx.doi.org/10.1016/j.gloenvcha.2006.01.001.

S. Stisen, A. L. Højberg, L. Troldborg, J. C. Refsgaard, B. S. B. Christensen, M. Olsen, and H. J. Henriksen. On the importance of appropriate precipitation gauge catch correction for hydrological modelling at mid to high latitudes. *Hydrol. Earth Syst. Sci.*, 16:4157–4176, 2012. doi: 10.5194/hess-16-4157-2012.

B. D. Stocker, R. Roth, F. Joos, R. Spahni, M. Steinacher, S. Zaehle, L. Bouwman, Xu-Ri, and I. C. Prentice. Multiple greenhouse-gas feedbacks from the land biosphere under future climate change scenarios. *Nature Climate Change*, 3: 666–672, 2013. doi: 10.1038/nclimate1864.

P. A. Stott. Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.*, 30:1728–1731, 2003a.

P. A. Stott. Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.*, 30:1724, 2003b. doi: 10.1029/2003GL017324.

K. Takahashi. Radiative constraints on the hydrological cycle in an idealized radiative convective equilibrium model. *J. Atmos. Sci.*, 66:77–91, 2009. doi: 10.1175/2008JAS2797.1.

K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, 93:485–498, 2012.

D. W. J. Thompson, J. M. Wallace, J. J. Kennedy, and P. D. Jones. An abrupt drop in Northern Hemisphere sea surface temperature around 1970. *Nature*, 467:444–447, 2010. doi: 10.1038/nature09394.

L. Thorpe and T. Andrews. The physical drivers of historical and 21st century global precipitation changes. *Environ. Res. Lett.*, 9:064024, 2014.

T. Toniazzo and A. A. Scaife. The influence of ENSO on winter North Atlantic climate. *Geophys. Res. Lett.*, 33:L24704, 2006. doi: 10.1029/2006GL027881.

K. E. Trenberth, A. Dai, R. M. Rasmussen, and D. B. Parsons. The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, 84:1205–1217, 2003. doi: 10.1175/BAMS-84-9-1205.

K. E. Trenberth, L. Smith, T. Qian, A. Dai, and J. Fasullo. Estimates of the global water budget and its annual cycle using observational and model data. *J. Hydrometeor.*, 8:758–769, 2007. doi: 10.1175/JHM600.1.

S. Twomey. The influence of pollution on the shortwave albedo of clouds. *J. Atmos. Sci.*, 34:1149–1152, 1977.

G. K. Vallis. *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press, Cambridge, U.K., 2006.

R. Vautard, J. Cattiaux, P. Yiou, J.-N. Thepaut, and P. Ciais. Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness. *Nature Geoscience*, 3:756–761, 2010. doi: 10.1038/ngeo979.

G. A. Vecchi, B. J. Soden, A. T. Wittenberg, I. M. Held, A. Leetmaa, and M. J. Harrison. Weakening of tropical Pacific atmospheric circulation due to anthropogenic forcing. *Nature*, 441:73–76, 2006. doi: 10.1038/nature04744.

C. J. Vörösmarty and B. Moore III. Modeling basin-scale hydrology in support of physical climate and global biogeochemical studies: An example using the Zambezi River. *Surv. Geophys.*, 12:271–311, 1991. doi: 10.1007/BF01903422.

C. J. Vörösmarty, K. Sharma, B. M. Fekete, A. H. Copeland, J. Holden, J. Marble, and J. A. Lough. The storage and aging of continental runoff in large reservoir systems of the world. *Ambio*, 26:210–219, 1997.

R. S. Vose, R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. K. Eischeid. The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. Rep. ORNL/CDIAC-53 NDP-041, Oak Ridge National Laboratory, 1992.

M. T. Walter, D. S. Wilks, J.-Y. Parlange, and R. L. Schneider. Increasing evapotranspiration from the conterminous united states. *J. Hydrometeor.*, 5:405–408, 2004.

H. Wan, X. Zhang, F. Zwiers, and S.-K. Min. Attributing northern high-latitude precipitation change over the period 1966-2005 to human influence. *Climate Dyn.*, pages 1–14, 2014. doi: 10.1007/s00382-014-2423-y.

R. A. Warren, D. J. Kirshbaum, R. S. Plant, and H. W. Lean. A Boscastle-type quasi-stationary convective system over the UK Southwest Peninsula. *Quart. J. Roy. Meteor. Soc.*, 140:240–257, 2014. doi: 10.1002/qj.2124.

F. C. S. Weiland, L. P. H. van Beek, J. C. J. Kwadijk, and M. F. P. Bierkens. Global patterns of change in discharge regimes for 2100. *Hydrol. Earth Syst. Sci.*, 16:1047–1062, 2012. doi: 10.5194/hess-16-1047-2012.

F. J. Wentz and M. Schabel. Precise climate monitoring using complementary satellite data sets. *Nature*, 403:414–416, 2000. doi: 10.1038/35000184.

F. J. Wentz, L. Ricciardulli, K. Hilburn, and C. Mears. How much more rain will global warming bring? *Science*, 317:233–235, 2007.

S. Westra, L. V. Alexander, and F. W. Zwiers. Global increasing trends in annual maximum daily precipitation. *J. Climate*, 26:3904–3918, 2013. doi: 10.1175/JCLI-D-12-00502.1.

L. J. Wilcox, E. J. Highwood, and N. J. Dunstone. The influence of anthropogenic aerosol on multi-decadal variations of historical global climate. *Environ. Res. Lett.*, 8:024033, 2013.

M. Wild. Enlightening global dimming and brightening. *Bull. Amer. Meteor. Soc.*, 93:27–37, 2012.

D. S. Wilks. *Statistical methods in the atmospheric sciences*. International Geophysics Series. Academic Press, 2nd edition, 2006.

K. M. Willett, P. D. Jones, N. P. Gillett, and P. W. Thorne. Recent changes in surface humidity: Development of the HadCRUH dataset. *J. Climate*, 21: 5364–5383, 2008. doi: 10.1175/2008JCLI2274.1.

P. Wu, N. Christidis, and P. Stott. Anthropogenic impact on Earth's hydrological cycle. *Nature Climate Change*, 3:807–810, 2013.

K. Xu, J. D. Milliman, and H. Xu. Temporal trend of precipitation and runoff in major chinese rivers since 1951. *Global Planet. Change*, 73:219–232, 2010. doi: 10.1016/j.gloplacha.2010.07.002.

D. Yang, B. E. Goodison, J. R. Metcalfe, V. S. Golubev, R. Bates, T. Pangburn, and C. L. Hanson. Accuracy of NWS 8-inch standard non-recording precipitation gauge: Results of WMO intercomparison,. *J. Atmos. Oceanic Technol.*, 15:54–68, 1998.

D. Yang, D. Kane, Z. Zhang, D. Legates, and B. Goodison. Bias corrections of long-term (1973-2004) daily precipitation data over the northern regions. *Geophys. Res. Lett.*, 32:L19501, 2005.

H. Yang, S. Piao, Z. Zeng, P. Ciais, Y. Yin, P. Friedlingstein, S. Sitch, A. Ahlström, M. Guimberteau, C. Huntingford, S. Levis, P. E. Levy, M. Huang, Y. Li, X. Li, M. R. Lomas, P. Peylin, B. Poulter, N. Viovy, S. Zaehle, N. Zeng, F. Zhao, and L. Wang. Multi-criteria evaluation of discharge simulation in Dynamic Global Vegetation Models. *J. Geophys. Res.*, 2015. doi: 10.1002/2015JD023129.

S. Zaehle and A. D. Friend. Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. *Global Biogeochem. Cycles*, 24:GB1005, 2010. doi: 10.1029/2009GB003521.

S. Zaehle, A. D. Friend, P. Friedlingstein, F. Dentener, P. Peylin, and M. Schulz. Carbon and nitrogen cycle dynamics in the O-CN land surface model: 2. Role of the nitrogen cycle in the historical terrestrial carbon balance. *Global Biogeochem. Cycles*, 24:GB1006, 2010. doi: 10.1029/2009GB003522.

L. Zhang, K. Hickel, W. R. Dawes, F. H. S. Chiew, A. W. Western, and P. R. Briggs. A rational function approach for estimating mean annual evapotranspiration. *Water Resour. Res.*, 40:W02502, 2004. doi: 10.1029/2003WR002710.

X. Zhang, F. W. Zwiers, G. C. Hegerl, F. H. Lambert, N. P. Gillett, S. Solomon, P. A. Stott, and T. Nozawa. Detection of human influence on twentieth-century precipitation trends. *Nature*, 448:461–465, 2007. doi: 10.1038/nature06025.

X. Zhang, H. Wan, F. W. Zwiers, G. C. Hegerl, and S.-K. Min. Attributing intensification of precipitation extremes to human influence. *Geophys. Res. Lett.*, 40:5252–5257, 2013. doi: 10.1002/grl.51010.

C. Zhao, X. Tie, and Y. Lin. A possible positive feedback of reduction of precipitation and increase in aerosols over eastern central China. *Geophys. Res. Lett.*, 33:L11814, 2006. doi: 10.1029/2006GL025959.