Editor

**General Comment**: Dear Authors, The paper is accepted once you have addressed all the following minor revisions required by the reviewers. I report them below. Please, provide a revised manuscript and, if needed a point-to-point answer to them.

**Response**: Thanks for your effort to handle our submission.

We have carefully checked all the comments and made the corresponding revisions into the revised paper. In particular, Following the Reviewer #2’s constructive comment, we divided the Section 3.4 into two parts for readability. These revisions made the revised paper clearer. Below please find our point to point response to your comments.
Reviewer #1

**General Comment:** The authors have taken a very substantive redraft following the review process and adequately accounted for the vast majority of reviewer comments. The revised text is very much clearer and organized more logically which greatly aids readability. It is, to my opinion, unfortunate that the authors largely left the figures as they stood originally as I still find these figures very hard to decipher. That, ultimately is the authors’ choice. I have a number of mainly minor comments detailed below.

**Response:** Thanks for your effort to evaluate our manuscript and high recommendation for the revised paper. As the other reviewer agree on the color use of the figures, we would like to keep their current form. Below please find our point to point response to your comments.

**Specific Comments:**

1) **Comment:** First, I wonder to what extent the results highlight inadequacy in reanalyses accounting for aerosol effects either via their omission, climatology, or even when applied (MERRA-2) if the direct and indirect effects are inadequately accounted for. It seems to me that a greater emphasis may be warranted in the discussion section in regard to whether the errors that appear to systematically affect several products may plausibly physically arise from inadequacies in the specification of and/or modelling of the impacts of aerosols. If the quantified errors are shown to be consistent with missing aerosol effects then that would be enormously helpful to point out more strongly to the reanalysis producers so they can perhaps undertake further experimentation and development work to rectify?

**Response:** Thanks for your comment. Because of various assimilated observations, different forecast models and different assimilated methods in different reanalysis products, it’s very difficult to quantify the presenting errors in regional warming from aerosol information errors in reanalyses. The result more advanced than previous studies is that “Comparing the $T_a$ values from MERRA2 and MERRA shows that MERRA2 displays improved performance over northern China, as reflected by an increase in the correlation coefficient of 0.1 and a reduction in the RMSE of 0.1°C (Fig. 3).”

So, following the revision of Reviewer #2, we revised the sentence in Lines 694-697 in Section Discussion:

MERRA2’s pioneering incorporation of time-varying aerosol loadings provides a way of improving the representation of regional temperature changes over regions such as the North China Plain where the impacts of aerosols on surface temperatures are significant.
2) **Comment:** Lines 236-238 this wording isn't quite right. I think you intend something like "thus the inclusion of ERA-20CM in this study provides a useful benchmark series against which to ascertain the skill that is added via assimilation of various observations" or something like that?

**Response:** Thanks for your providing such information. Corrected as suggested: thus, the inclusion of ERA-20CM in this study provides a useful benchmark series against which to ascertain the skill that is added by assimilating various observations and to cognize the advantage of ensemble simulations.

3) **Comment:** Line 273 impact of data inhomogeneities on (inhomogeneities and not homogenization here).

**Response:** Corrected as suggested.

4) **Comment:** Line 426 Fig. 4 (Fig. to Figs.)

**Response:** Corrected as suggested.
Reviewer #2

General Comment: I have read through the revised version of this paper. I thank the authors for taking note and responding to comments made in my original review. I recommend this version for publication, but do have the following specific comments on this version of the paper.

Response: Thanks for your high recommendation of our submission. Below please find our point to point response to your comments.

Specific Comments:

1) Comment: The Abstract and Conclusions identify welcome developmental steps involving incorporation of aerosol information and use of ensemble techniques. As pointed out in the body of the paper, the representations of \(T_a\) in ERA-Interim and JRA-55 are much better than those in other reanalyses, almost certainly in significant part due to their analysis of synoptic observations of the variable. This could be mentioned in the Abstract and Conclusions as another way the near-surface products of other reanalyses might be improved.

Response: Thanks for your providing such information, which is added in Abstract and Conclusions: In addition, the analysis of \(T_a\) observations helps representing regional warming in ERA-Interim and JRA-55.

2) Comment: Table 1 has been tidied up following a comment in my original review, but the entry for ERA-20CM remains incorrect. The entry under "Assimilation System" for ERA-20CM has been changed from 4D-Var to 3D-Var, but this is still not right. There should be no entry under “Assimilation System” for ERA-20CM as it did not directly assimilate any atmospheric observations. Observations influence ERA-20CM only indirectly, through the forcing provided by CMIP5, and the prescribed SST, sea-ice and other model fields. ERA-20CM had no assimilation system of its own.

Response: We agree with the reviewer and deleted the “3D-Var” in the row of ERA-20CM.

3) Comment: Regarding point (6) of my original review, I can accept the terminology “NWP-like reanalysis” and “climate reanalysis”, even though I still do not particularly
like it. In Dee et al. (2014) – of which I am a co-author – we use the phrases “NWP-like reanalysis” and “extended climate reanalysis” which is subtly different, in that we (or at least I) regard the NWP-like reanalysis as one type of climate reanalysis and the extended climate reanalysis as another (longer) type. However, this slightly different nomenclature will not work in the present authors’ case as CFSR, which they classify as a climate analysis, is not “extended” in the sense used by Dee et al.

Response: Thanks for your comment. We provided an explanation in Lines 254-256: Note that the CFSR is classified into climate reanalysis in this study, mainly because it adopts a climate forecast system (Table 1).

4) Comment: Line 229. It would probably be better to write “incorporate many observations” rather than “incorporate some observations”, even though the reanalyses have access to under 10% of the Chinese observations available to the authors. Globally, the number of observations used by ERA-Interim is some thirty thousand per day for the 1980s, and that number has risen considerably in recent years.

Response: Corrected as suggested.

5) Comment: Lines 310 and 311. The sentence that spans these lines needs rephrasing as the correlation coefficient does not assess the absolute value.

Response: Corrected as “The bias, root mean squared error (RMSE) and standard deviation are used to assess the absolute value of $T_a$.”

6) Comment: Line 359. “A homogeneous adjustment” should, I believe, be changed to “a homogenizing adjustment”.

Response: Corrected as suggested.


Response: Corrected as suggested.
Comment: Line 431. It would be better to replace “jointly determined” by “determined in part”, “determined to a significant extent” or some such phrasing. This is needed because surface air temperature is sensitive to a number of factors – such as circulation patterns and snow cover – that are not considered in the authors’ analysis, or only partially considered – a circulation anomaly may change Ta through differences in precipitation frequency that are taken into account, but may also change temperature due to advection. Similar qualifications are needed in several similar statements made in this and the following section.

Response: Following the reviewer’s suggestion, we replaced replace “jointly determined” by “determined in part”.

9)

Comment: Line 512. The trend in precipitation frequency is given in units of days/decade, implying that the precipitation frequency has a unit of days. How is precipitation frequency defined? Is it the number of days per month with appreciable rainfall? Maybe this is stated somewhere in the paper, but I do not recall seeing it.

Response: This information was revised in Lines 216-219: Additionally, precipitation frequency is defined as days in a year with daily precipitation at least 0.1 mm in this study, which has been shown to provide a good indication of the effects of precipitation on the interannual variability and trends in $T_a$ (Zhou et al., 2017).

10)

Comment: Page 527 and 528. Elsewhere the biases in temperature trends are related to biases in precipitation frequency, Ld etc.. But in this sentence it is the other way round. Is this a mistake?

Response: Yes, we have corrected it.

11)

Comment: Section 3.4 is a long one, and a more succinct synthesis of the results would probably help many readers.

Response: Thanks for your suggestions. We divided the original Section 3.4 into two parts, i.e., 1) The whole of China and 2) Seven Subregions. Accordingly, we adjusted the structure of this section for good readability.
Comment: Line 578. Incorporating observations of Ta as done by ERA-Interim and JRA-55 is expected generally to reduce biases in the trends of analyses of Ta, not to introduce biases, as the reanalyses tend generally to reproduce the trends present in the observations, even in the presence of biases in the assimilating models. An exception can occur for regions where observations are not available for a substantial part of the time, particularly early or late in the period covered by the reanalysis. In this case, trends will be in error if the assimilating model has significant biases that are corrected at times when there are observations, but uncorrected when observations are absent.

Response: Thanks for your comment. We delete this sentence.

13)

Comment: Lines 617 and 618. Elevation differences can occur over time if stations are relocated, but this should be largely taken care of by the homogenization.

Response: Thanks for your comment. We delete this sentence.

14)

Comment: Lines 647 and 648. ERA-20CM should be referred to as ensemble simulation not ensemble forecasting.

Response: Corrected as suggested.

15)

Comment: Lines 663-670. It is stated that the comparison is for the reanalyses that do not incorporate observations, but ERA-Interim and JRA-55 are listed in the comparison. This needs amending.

Response: Thanks for your comment. We revised it in Lines 663-666:

To provide a preliminary discussion of the improvements in climate forecast models in reflecting patterns in climate trends, we compare the spatial patterns of the biases in the trends in $R_s$, precipitation frequency and $L_d$ because observations of these variables are not included in the reanalyses.

16)

Comment: Lines 691 and 691. This reads rather as if MERRA2 included time-vary aerosol loading in order to make a pioneering attempt to improve regional warming over the North China Plain. One suspects that the motivation for including time-varying aerosols in MERRA2 was much more general than this. The text could
be amended slightly to read something like “MERRA2’s pioneering incorporation of time-varying aerosol loadings provides a way of improving the representation of regional temperature changes over regions such as the North China Plain where the impacts of aerosols on surface temperatures are significant.”

Response: Thanks for your revision, which is added in Lines 694-697:

MERRA2’s pioneering incorporation of time-varying aerosol loadings provides a way of improving the representation of regional temperature changes over regions such as the North China Plain where the impacts of aerosols on surface temperatures are significant.
On the Suitability of Current Atmospheric Reanalyses for Regional Warming Studies over China

Chunlue Zhou¹, Yanyi He¹, Kaicun Wang*¹

¹College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China

*Corresponding Author: Kaicun Wang, College of Global Change and Earth System Science, Beijing Normal University. Email: kcwang@bnu.edu.cn; tel.: +86 (10)-58803143; fax: +86 (10)-58800059.

Submitted to Atmospheric Chemistry and Physics

March 22, May 12, 2018
Reanalyses are widely used because they add value to routine observations by generating physically or dynamically consistent and spatiotemporally complete atmospheric fields. Existing studies include extensive discussions of the temporal suitability of reanalyses in studies of global change. This study adds to this existing work by investigating the suitability of reanalyses in studies of regional climate change, in which land-atmosphere interactions play a comparatively important role. In this study, surface air temperatures ($T_a$) from 12 current reanalysis products are investigated; in particular, the spatial patterns of trends in $T_a$ are examined using homogenized measurements of $T_a$ made at ~2200 meteorological stations in China from 1979 to 2010. The results show that ~80% of the mean differences in $T_a$ between the reanalyses and the in situ observations can be attributed to the differences in elevation between the stations and the model grids. Thus, the $T_a$ climatologies display good skill, and these findings rebut previous reports of biases in $T_a$. However, the biases in the $T_a$ trends in the reanalyses diverge spatially (standard deviation=0.15-0.30°C/decade using 1°×1° grid cells). The simulated biases in the trends in $T_a$ correlate well with those of precipitation frequency, surface incident solar radiation ($R_s$), and atmospheric downward longwave radiation ($L_d$) among the reanalyses ($r$=-0.83, 0.80 and 0.77; $p<0.1$) when the spatial patterns of these variables are considered. The biases in the trends in $T_a$ over southern China (on the order of -0.07°C/decade) are caused by biases in the trends in $R_s$, $L_d$ and precipitation frequency on the order of 0.10°C/decade, -0.08°C/decade, and -0.06°C/decade,
respectively. The biases in the trends in $T_a$ over northern China (on the order of 
-0.12°C/decade) result jointly from those in $L_d$ and precipitation frequency. Therefore,
improving the simulation of precipitation frequency and $R_s$ helps to maximize the 
signal component corresponding to regional climate. In addition, the analysis of $T_a$
observations helps representing regional warming in ERA-Interim and JRA-55.

Incorporating vegetation dynamics in reanalyses and the use of accurate aerosol
information, as in the Modern-Era Retrospective Analysis for Research and
Applications, version 2 (MERRA-2), would lead to improvements in the modelling of
regional warming. The use of the ensemble technique adopted in the
twentieth-century atmospheric model ensemble ERA-20CM significantly narrows the
uncertainties associated with regional warming in reanalyses (standard
deviation=0.15°C/decade).
1. Introduction

Observations and models are two fundamental approaches used in the understanding of climate change. Observations provide a direct link to the climate system via instruments, whereas models provide an indirect link and include information derived from measurements, prior knowledge and theory.

A large number of meteorological observations have been accumulated. These measurements, which are derived from a variety of sources, such as surface stations, ships, buoys, radiosondes, airplanes and satellites, record quantities that include near-surface and upper-air temperatures, humidity, wind and pressure. They constitute a major source of atmospheric information through the depth of the troposphere but suffer from incomplete spatiotemporal coverage and observation errors, including systematic, random and representation errors. Recent satellite-based observations have much better coverage; however, they suffer from other notable limitations, including temporal inhomogeneities (e.g., satellite drift) and retrieval errors (Bengtsson et al., 2007). These spatiotemporally varying gaps restrict the effective application of observations alone in climate research.

To fill in the gaps in observations, models are needed. Such models can be very simple; examples of simple models include linear interpolation or geo-statistical approaches that are based on the spatial and temporal autocorrelation of the observations. However, these models lack the necessary dynamical or physical mechanisms. Given the steady progress of numerical weather prediction (NWP) models in characterizing the global atmospheric circulation in the early 1980s (Bauer...
et al., 2015), the first generation of reanalyses was produced by combining observations and dynamic models to provide the first global atmospheric datasets for use in scientific research (Bengtsson et al., 1982a, b).

After realizing the great value of this kind of reanalysis in atmospheric research, a step forward was taken with the suggestion made by Bengtsson and Shukla (1988) and Trenberth and Olson (1988) that most meteorological observations should be optimally assimilated under a fixed dynamical system over a period of time long enough to be useful for climate studies. In this way, available observations are ingested by advanced data assimilation techniques to provide a continuous initial state for an NWP model to produce the next short-term forecast. This procedure thus generates physically consistent and spatiotemporally complete three-dimensional atmospheric fields that are updated in light of observations.

Taking this suggestion as a guide, and given the improvements that have been made since the mid-1990s in the integrity of the observations, the models and the assimilation methods used, successive generations of atmospheric reanalyses established by several institutes have improved in quality. These reanalyses include the first two generations of global reanalyses produced by the National Centers for Environmental Prediction, NCEP-R1 (Kalnay et al., 1996) and NCEP-R2 (Kanamitsu et al., 2002) and the reanalyses produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), ERA-15 (Gibson et al., 1997), ERA-40 (Uppala et al., 2005), and ERA-Interim (Dee et al., 2011b); the Japanese Meteorological Agency, JRA-25 (Onogi et al., 2007) and JRA-55 (Kobayashi et al., 2015); and the National
Aeronautics and Space Administration, the Modern-Era Retrospective Analysis for Research and Applications (MERRA) (Rienecker et al., 2011) and its updated version, MERRA-2 (Reichle et al., 2017).

These reanalyses produce global gridded datasets that cover multiple time scales and include a large variety of atmospheric, oceanic and land surface parameters, many of which are not easily or routinely observed but are dynamically constrained by large numbers of observations from multiple sources assimilated using fixed NWP models. During the data assimilation, prior information on uncertainties in the observations and models are used to perform quality checks, to derive bias adjustments and to assign proportional weights. Therefore, such reanalyses add value to the instrumental record through their inclusion of bias adjustments, their broadened spatiotemporal coverage and their increased dynamical integrity or consistency.

Previous studies have revealed that such reanalyses have contributed significantly to a more detailed and comprehensive understanding of the dynamics of the Earth’s atmosphere (Dee et al., 2011b; Kalnay et al., 1996; Nguyen et al., 2013; Kidston et al., 2010; Simmonds and Keay, 2000; Simmons et al., 2010; Mitas and Clement, 2006). Extensive assessment studies have reported that most reanalyses display a certain level of performance in terms of their absolute values (Betts et al., 1996; Zhou and Wang, 2016b; Betts et al., 1998), interannual variability (Lin et al., 2014; Lindsay et al., 2014; Zhou and Wang, 2017a, 2016a; Wang and Zeng, 2012), distributions (Gervais et al., 2014; Heng et al., 2014; Mao et al., 2010) and relationships among variables (Niznik and Lintner, 2013; Cash et al., 2015; Zhou et al., 2017; Zhou and Wang,
2016b; Betts, 2004) over regions worldwide. However, these aspects of reanalyses still contain certain errors that restrict the general use of reanalyses, especially in climate applications.

The errors displayed by reanalysis products arise from three sources: observation error, model error and assimilation error (Thorne and Vose, 2010; Parker, 2016; Lahoz and Schneider, 2014; Dee et al., 2014; Zhou et al., 2017). Specifically, observation error incorporates systematic and random errors in instruments and their replacements, errors in data reprocessing and representation error, which arises due to the spatiotemporal incompleteness of observations (Dee and Uppala, 2009; Desroziers et al., 2005). Model error refers mainly to the inadequate representation of physical processes in NWP models (Peña and Toth, 2014; Bengtsson et al., 2007), such as the lack of time-varying surface conditions, such as vegetation growth (Zhou and Wang, 2016b; Trigo et al., 2015), and incomplete cloud-precipitation-radiation parameterizations (Fujiwara et al., 2017; Dolinar et al., 2016). Assimilation error describes errors that arise in the mapping of the model space to the observation space and errors in the topologies of cost functions (Dee, 2005; Dee and Da Silva, 1998; Lahoz and Schneider, 2014; Parker, 2016).

These reanalyses mentioned above consist of the true climate signal and the nonlinear interactions among the observation error, the model error, and the assimilation error that arise during the assimilation process. These time-varying errors can introduce spurious trends without being eliminated by data assimilation systems. Many spurious variations in climate signals were also identified in the
early-generation reanalyses (Bengtsson et al., 2004; Andersson et al., 2005; Chen et al., 2008; Zhou and Wang, 2016a, 2017a; Zhou et al., 2017; Schoeberl et al., 2012; Xu and Powell, 2011; Hines et al., 2000; Cornes and Jones, 2013). Therefore, reanalyses produced using the existing reanalysis strategy may not accurately capture climate trends (Trenberth et al., 2008), even though they may contain relatively accurate estimates of synoptic or interannual variations in the Earth’s atmosphere.

An emerging requirement for climate applications of reanalysis data is the accurate representation of decadal variability, further increasing the confidence in the estimation of climate trends. This kind of climate reanalysis is required to be free, to a great extent, from other spurious non-climatic signals introduced by changing observations, model imperfections and assimilation error; that is, they must maintain temporal consistency. Therefore, the extent to which climate trends can be assessed using reanalyses attracts much attention and sparks heated debates (Thorne and Vose, 2010; Dee et al., 2011a; Dee et al., 2014; Bengtsson et al., 2007).

Given the great progress that has been made in climate forecasting models (which provide more accurate representations of climate change and variability) and coupled data assimilation, many efforts have been made by several institutes to build consistent climate reanalyses using the strategy of assimilating a relatively small number of high-quality long-term observational datasets. The climate reanalyses of this new generation extend back to the late nineteenth century and include the Climate Forecast System Reanalysis (CFSR), which is produced by the National Centers for Environmental Prediction (Saha et al., 2010); NOAA 20CRv2c, which is produced by
the University of Colorado’s Cooperative Institute for Research in Environmental Sciences (CIRES) in cooperation with the National Oceanic and Atmospheric Agency (NOAA) (Compo et al., 2011); and ERA-20C (Poli et al., 2016), ERA-20CM (Hersbach et al., 2015) and CERA-20C (Laloyaux et al., 2016), which are produced by the ECMWF. Compo et al. (2013) suggested that the NOAA 20CRv2c reanalysis can reproduce the trend in global mean surface air temperatures. In addition, the uncertainties estimated from multiple ensembles are provided to increase the confidence of the climate trends (Thorne and Vose, 2010; Dee et al., 2014).

From NWP-like reanalyses to climate reanalyses, existing studies focus mainly on comparing the differences in temporal variability between the reanalyses and observations using some statistical metrics, e.g., the mean values, standard deviations, interannual correlations, probability density functions and trends of surface air temperature over regions worldwide. These evaluations provide insight into the temporal evolution of the Earth’s atmosphere. However, they lack the performance evaluations used in reanalyses in representing the spatial patterns of these statistics associated with the role of the coupled land-atmosphere and dynamical processes of the climate system. Moreover, the assessment of these spatial patterns provides a direct means of examining the most prominent advantage of reanalyses over geo-statistical interpolation; thus, the spatial patterns require comprehensive investigation.

This study employs high-density station-based datasets of quantities including surface air temperatures ($T_a$), the surface incident solar radiation ($R_s$), the surface
downward longwave radiation \((L_d)\), and precipitation measured at \(~2200\) meteorological stations within China from 1979 to 2010. It provides a quantitative examination of the simulated patterns of variations in \(T_a\) in both the NWP-like and climate reanalyses and considers the climatology, the interannual variability, the mutual relationships among relevant quantities, the long-term trends and their controlling factors. The results indicate the strengths and weaknesses of the current reanalyses when applied in regional climate change studies and provide possible ways to improve these reanalyses in the near future.

2. Data and Methods

2.1 Observational Datasets

The latest comprehensive daily dataset (which contains averages at 0, 6, 12, and 18 UTC) of quantities that include \(T_a\), precipitation, sunshine duration, relative humidity, water vapor pressure, surface pressure and the cloud fraction from approximately 2400 meteorological stations in China from 1961 to 2014, of which only approximately 194 participate in global exchanges, is obtained from the China Meteorological Administration (CMA; [http://data.cma.cn/data](http://data.cma.cn/data)). Approximately 2200 stations with complete and homogeneous data are selected for use in this study (Wang and Feng, 2013; Wang, 2008; Wang et al., 2007). The high density of meteorological stations in China promotes the representation of regional patterns in surface warming by reanalyses and the assessment of the skill of simulations. \(R_s\) values based on the revised Ångström-Prescott equation (Wang et al.,
2015; Yang et al., 2006; Wang, 2014) are used in this study. The derived $R_s$ values consider the effects of Rayleigh scattering, water vapor absorption and ozone absorption (Wang et al., 2015; Yang et al., 2006) and can accurately reflect the effects of aerosols and clouds on $R_s$ over China (Wang et al., 2012; Tang et al., 2011). Several intensive studies have reported that the derived $R_s$ values can accurately depict the interannual, decadal and long-term variations in $R_s$ (Wang et al., 2015; Wang, 2014; Wang et al., 2012).

$L_d$ is typically estimated by first determining the clear-sky radiation and atmospheric emissivity (Brunt, 1932; Choi et al., 2008; Bilbao and De Miguel, 2007), and then correcting for the cloud fraction (Wang and Liang, 2009; Wang and Dickinson, 2013). The derived $L_d$ values can directly reflect the greenhouse effect of atmospheric water vapor and clouds. Additionally, the precipitation frequency is defined as days in a year with daily precipitation of at least 0.1 mm in this study, which has been shown to provide a good indication of the effects of precipitation on the interannual variability and trends in $T_a$ (Zhou et al., 2017). Taken together, the derived $R_s$ and $L_d$ values are able to physically quantify the effects of solar radiation and the greenhouse effect on surface warming. Precipitation frequency can regulate the partitioning of available energy into latent and sensible heat fluxes and thus modulates the variations in $T_a$ (Zhou et al., 2017; Zhou and Wang, 2017a).

2.2 Reanalysis Products

All of the major global atmospheric reanalysis products are included in this study (Table 1). The reanalyses are summarized below in terms of three aspects, i.e., the
observations assimilated and the forecast model and assimilation method used. The
NWP-like reanalyses assimilate many conventional and satellite datasets from
multiple sources (Table 1) to characterize the basic upper-air atmospheric fields; the
spatiotemporal errors of these datasets vary with time. In particular, the ERA-Interim
and JRA-55 reanalyses incorporate some many observations of \( T_a \), and the MERRA2
reanalysis includes aerosol optical depth estimates from satellite retrievals and model
simulations based on emission inventories, whereas most of the other reanalyses use
climatological aerosols (Table 1). To derive consistent long-term climate signals, the
new strategy adopted by climate reanalyses involves the assimilation of a small
number of relatively effective observed variables, e.g., surface pressure (Table 1).

Except for its lack of the assimilation of surface pressure, ERA-20CM employs the
same forecast model and external forcings as ERA-20C (Table 1); thus, the inclusion
of ERA-20CM in this study will provide insight into the suitability of current
atmospheric reanalyses in studies of regional warming thus, the inclusion of
ERA-20CM in this study provides a useful benchmark series against which to
ascertain the skill that is added by assimilating various observations and to cognize
the advantage of ensemble simulations. The reanalyses adopt different sea surface
temperatures (SSTs) and sea ice concentrations for different time periods, which may
lead to temporal discontinuities in the climate signals derived from the reanalyses
(Table 1). To address this issue, the boundary conditions in CFSR are derived from its
coupled ocean-sea ice models instead of observations (Table 1). CFSR, NOAA
20CRv2c and NOAA 20CRv2 use monthly greenhouse gases (GHGs) with annual
means near those used in CMIP5. On the other hand, in ERA-Interim, the GHGs
increase more slowly than in CMIP5 after 2000. Finally, NCEP-R1 and NCEP-R2
adopt constant global mean concentrations of the GHGs (Table 1).

The forecast model is a fundamental component of a reanalysis that provides the
background fields to the assimilation system. Different reanalyses produced by a
single institute generally use similar physical parameterizations; however, updated
versions of these parameterizations and higher spatial resolutions are used in the
newer generations of these realizations (Table 1). Note that the CFSR is classified into
climate reanalysis in this study, mainly because it adopts a climate forecast system
(Table 1). The assimilation methods adopted by the current reanalyses incorporate
variational methods (3D-Var and 4D-Var) and the ensemble Kalman filter (EnKF)
approach (Table 1).

The 2-m $T_a$ in NCEP-1, NCEP-2, MERRA, MERRA-2, ERA-20C, ERA-20CM,
CERA-20C, NOAA 20CRv2c, NOAA 20CRv2 and CFSR are model-derived fields
that are functions of the surface skin temperature, the temperature at the lowest model
level, the vertical stability and the surface roughness, which are constrained primarily
by observations of upper-air variables and the surface pressure (Kanamitsu et al.,
2002;Rienecker et al., 2011;Reichle et al., 2017;Poli et al., 2016;Hersbach et al.,
2015;Laloyaux et al., 2016;Compo et al., 2011;Saha et al., 2010). However, the $T_a$ in
ERA-Interim and JRA-55 are post-processing products by a relatively simple analysis
scheme between the lowest model level and the surface and are analysed using
ground-based observations of $T_a$, with the help of Monin-Obukhov similarity profiles
consistent with the model’s parameterization of the surface layer (Dee et al., 2011b; Kobayashi et al., 2015). Additionally, radiation calculations are diagnostically determined from the prognostic cloud condensate microphysics parameterization, and the cloud macrophysics parameterization assumes a maximum-random cloud overlapping scheme (Saha et al., 2010; Dolinar et al., 2016).

2.3 Method Used to Homogenize the Observed Time Series

Problems related to the observational infrastructure (e.g., instrument ageing and changes in observing practices) and station relocations can also lead to false temporal heterogeneity in time series. Therefore, it is necessary to diminish the impact of data inhomogeneities on the trends in the observed variables during the study period of 1979-2010.

We use the RHtestsV4 software package (Wang and Feng, 2013) to detect and homogenize the breakpoints in the monthly time series. The package includes two algorithms. Specifically, the PMFred algorithm is based on the penalized maximal $F$-test ($PMF$) without a reference series (Wang, 2008), and the PMTred algorithm is based on the penalized maximal $t$-test ($PMT$) with a reference series (Wang et al., 2007).

In this study, we first use the PMFred algorithm to identify potential reference series at the 95% significance level. We then reconstruct homogenous series for each inhomogeneous series using the following steps: 1) horizontal and vertical distances from the inhomogeneous station of less than 110 km and 500 m, respectively, are specified; 2) correlation coefficients between the first-order difference in the
homogeneous series with that in the inhomogeneous one exceeding 0.9 are required; 
and 3) the first ten homogeneous series are averaged using inverse distance weighting 
to produce a reference series for the inhomogeneous station. Finally, we apply the 
PMTred algorithm to test all of the inhomogeneous series using the nearby reference 
series. Several intensive studies have been conducted that indicate the PMTred 
algorithm displays good performance in detecting change points in inhomogeneous 
series (Venema et al., 2012; Wang et al., 2007).

If a breakpoint is found to be statistically significant, the quantile-matching (QM) 
adjustment in RHtestsV4 is recommended for making adjustments to the time series 
(Wang et al., 2010; Wang and Feng, 2013); in such cases, the longest available 
segment from 1979 to 2010 is used as the base segment. The QM adjustment aims to 
match the empirical distributions from all of the detrended segments with that of the 
specific base segment (Wang et al., 2010). In addition, we replicate the procedures 
above for the sparsely distributed stations over western China and the Tibetan Plateau. 
The PMTred algorithm and the QM adjustment have recently been used successfully 
to homogenize climatic time series (Aarnes et al., 2015; Tsidu, 2012; Dai et al., 
2011; Siswanto et al., 2015; Wang and Wang, 2016; Zhou et al., 2017).

As such, the significant breakpoints are detected and adjusted at a confidence 
level of 95% at 1092 of the 2193 (49.8%) stations for the $T_a$ time series; 1079 of the 
2193 (49.2%) stations for the $R_s$ time series; 64 of the 2193 (2.9%) stations for 
precipitation frequency time series; 971 of the 2193 (44.2%) stations for the $L_d$ time 
series; 944 of the 2193 (43.0%) stations for the water vapor pressure time series; and
956 of the 2193 (43.6%) stations for the cloud fraction time series.

2.4 Trend Calculations, Partial Linear Regression, and Total Least Squares

The bias, root mean squared error (RMSE) and standard deviation and correlation coefficient ($r$) are used to assess the absolute value of $T_a$. The trends in $T_a$ and the relevant variables are calculated using the ordinary least squares method (OLS) and the two-tailed Student’s $t$-test. To determine whether the reanalyses contain biases in these trends, the two-tailed Student’s $t$-test is also applied to the differences in the time series between the reanalyses and the homogeneous observations.

The partial least squares approach is used to investigate the net relationship between the detrended $T_a$ values and the relevant variables ($R_s$, $L_d$ and precipitation frequency) after statistically excluding the confounding effects among the relevant variables (Zhou et al., 2017). To evaluate the potential collinearity of independent variables in the regression model, the variance inflation factor ($VIF$) is calculated. The $VIF$s for $R_s$, precipitation frequency and $L_d$ are less than 4. Specifically, the $VIF$ for China of 2.19 is much less than the threshold of 10, above which the collinearity of regression models is bound to adversely affect the regression results (Ryan, 2008).

The Pearson correlation coefficient ($r$) is used to reveal the spatial relationship between $T_a$ and the relevant variables. To further investigate the relationship between the spatial distributions of the biases in the trends in $T_a$ and the relevant parameters among the twelve reanalysis products, the weighted total least squares (WTLS) is adopted, in which the spatial standard deviations and correlations of pairs of variables on $1^\circ \times 1^\circ$ grid cells are included (Reed, 1989; York et al., 2004; Golub and Van Loan,
\[\omega(x_i) = \frac{1}{\sigma_{x_i}^2}\]  
\[\omega(y_i) = \frac{1}{\sigma_{y_i}^2}\]  
\[W_i = \frac{\omega(x_i) \cdot \omega(y_i)}{\omega(x_i) + b^2 \omega(y_i) - 2b \cdot r_i \sqrt{\omega(x_i) \cdot \omega(y_i)}}\]  
\[U_i = x_i - \frac{\sum_{i=1}^{n} (W_i \cdot x_i)}{\sum_{i=1}^{n} (W_i)}\]  
\[V_i = y_i - \frac{\sum_{i=1}^{n} (W_i \cdot y_i)}{\sum_{i=1}^{n} (W_i)}\]  
\[\beta_i = W_i \left[ \frac{U_i}{\omega(y_i)} + \frac{b \cdot V_i}{\omega(x_i)} \right] \left( b \cdot U_i + V_i \right) \frac{r_i}{\sqrt{\omega(x_i) \cdot \omega(y_i)}}\]  
\[b = \frac{\sum_{i=1}^{n} W_i \cdot \beta_i \cdot V_i}{\sum_{i=1}^{n} W_i \cdot \beta_i \cdot U_i}\]

where \(x_i\) and \(y_i\) are the median trends in \(x\) and \(y\) (e.g., \(T_a\) and \(R_s\)) for the \(i^{th}\) reanalysis product; \(\sigma_{x_i}\), \(\sigma_{y_i}\) and \(r_i\) are the spatial standard deviations and correlations of the trends in \(x\) and \(y\) for the \(i^{th}\) reanalysis product; \(\beta_i\) is the least squares-adjusted value; \(W_i\) is the weight of the residual error; and \(b\) is the slope estimated by iterative methods with a relative tolerance of \(10^{-16}\).

The Monte Carlo method with 10000 experiments is applied to estimate the 90% confidence intervals of the slope \(b\). In the Monte Carlo method, the grid index for the \(1^\circ \times 1^\circ\) grid cells over China, which ranges from 1 to 691, is generated as a random number. On this basis, we can sample the spatial pattern in the biases in the trends in \(T_a\), \(R_s\), \(L_d\) and precipitation frequency. We then calculate the median trends and their spatial standard deviations and correlations for each experiment used in the WTLS.
3. Results

3.1 Dependency of Surface Air Temperature Differences on Elevation Differences

Fig. 1 illustrates the differences in $T_a$ from the NWP-like reanalyses and climate reanalyses relative to the homogenized station-based observations over China during the period of 1979-2010. When the $T_a$ values measured at the stations are compared directly with those in the corresponding model grid cells, the results indicate that the reanalysis products underestimate $T_a$ over most of the regions in China (by -0.28°C to -2.56°C). These discrepancies are especially pronounced over the Tibetan Plateau and Middle China, where the underestimation ranges from -2.75°C to -7.00°C and from -1.19°C to -2.91°C, respectively (Fig. 1 and Table 2). A homogenizing adjustment of 0.03°C from the raw $T_a$ observations is insufficient to cancel the underestimation of $T_a$ by the reanalyses (Fig. 1 and Table 2). Similar biases in $T_a$ within various regions worldwide have been widely reported by previous studies (Mao et al., 2010; Pitman and Perkins, 2009; Reuten et al., 2011; Wang and Zeng, 2012; Zhou et al., 2017; Zhou and Wang, 2016b).

However, we found that the spatial patterns in the differences in $T_a$ are well correlated with the elevation differences between models and stations, as reflected by correlation coefficients ($r$) of 0.85 to 0.94 (Figs. 2 and S1). These results are in accordance with the reports from NCEP-R1, NCEP-R2 and ERA-40 (You et al., 2010; Ma et al., 2008; Zhao et al., 2008). The elevation differences ($\Delta$Height; Figs. 2 and S1) between the stations and the model grids consists of the filtering error in the
elevations used in the spectral models ($\Delta f$) and differences in the site-to-grid elevations ($\Delta s$) due to the complexity of the orographic topography. We further quantify the relative contributions of these factors to the $T_a$ differences. The elevation differences can explain approximately 80% of the $T_a$ differences; approximately 74% is produced by the site-to-grid elevation differences, and approximately 6% is produced by the filtering error in the elevations used in the spectral models (Fig. 2).

The regression coefficient of the differences in $T_a$ is approximately 6°C/1 km, which is similar to the lapse rate at the surface (Fig. 2). Lapse rate values that exceed 6°C/1 km can be seen over the Tibetan Plateau (shown as red dots in Fig. 2). This result is very consistent with the reported lapse rates over China (Li et al., 2015; Fang and Yoda, 1988). In addition, the rate of decrease in the model filtering error is approximately 4°C/1 km among the twelve reanalyses (Fig. 2). These results have important implications for the skill of the simulated $T_a$ climatologies of the twelve reanalyses over China.

### 3.2 Comparison of Regional-scale Surface Air Temperature Series

Fig. 3 shows Taylor diagrams of annual $T_a$ anomalies from the observations and reanalyses over China and its seven subregions. We find that the correlations between the annual $T_a$ anomalies in the twelve reanalysis products and the observations are reasonably strong, as reflected by a median $r$ of 0.95 (Fig. 3), despite the relatively weak correlations over the Tibetan Plateau associated with NCEP-R2 ($r=0.24$) and CFSR ($r=0.53$). The simulated time series of $T_a$ anomalies over eastern China are depicted most accurately by the reanalyses (Fig. 3c-g).
Overall, the NWP-like reanalyses (denoted by numbers 3-7) display better skill than the climate reanalyses (denoted by numbers 8-14) in this regard (Fig. 3). ERA-Interim and JRA-55 display the best performance in the simulated time series of $T_a$ anomalies over China ($r=1.00$, RMSE=0.05°C) and the seven regions ($r=0.98$, RMSE=0.1°C) (Fig. 3), perhaps due to their analysis of surface air temperature observations in ERA-Interim and JRA-55 (Table 1).

Comparing the $T_a$ values from MERRA2 and MERRA shows that MERRA2 displays improved performance over northern China, as reflected by an increase in the correlation coefficient of 0.1 and a reduction in the RMSE of 0.1°C (Fig. 3). This result may occur because MERRA2 includes time-varying aerosol loadings (Balsamo et al., 2015; Reichle et al., 2011). However, the incorporation of this information does not improve the results over Southeast China (Fig. 3h).

CERA-20C displays better performance than ERA-20C and ERA-20CM, perhaps related to the inclusion of coupled climate forecast models and data assimilation, as well as the assimilation of surface pressure data in CERA-20C (Fig. 3 and Table 1). NOAA 20CRv2c and NOAA 20CRv2 display moderate performance in this regard ($r=0.8$, RMSE=0.3°C) (Fig. 3), and the former reanalysis displays no improvement in performance, despite its use of new boundary conditions (Compo et al., 2011).

### 3.3 Key Factors Regulating Regional Temperature Change

This section discusses key factors that control regional temperature change from the perspective of energy balance and its partitioning. The $R_s$ heats the surface, and the portion of this radiation that becomes the sensible heat flux heats the air near the
Part of the energy absorbed by the surface is released back to space as outgoing longwave radiation; some of this radiation is reflected by clouds and is influenced by atmospheric water vapor, further warming the near-surface air (Wang and Dickinson, 2013). This process is known as the greenhouse effect on $T_a$ and is quantified by $L_d$.

Existing studies have suggested that precipitation frequency better represents the interannual variability in soil moisture in China than the precipitation amount (Wu et al., 2012; Piao et al., 2009; Zhou et al., 2017; Zhou and Wang, 2017a); in turn, soil moisture affects vegetation growth and drives changes in surface characteristics (e.g., surface albedo and roughness). These changes alter the partitioning of available energy and thus regulate changes in $T_a$.

Figs. 4 illustrates the partial relationships between the annual anomalies in $T_a$ and $R_s$, the precipitation frequency and $L_d$. The results show that $T_a$ is consistently positively correlated with $R_s$ (except over the Tibetan Plateau) and $L_d$; however, it is consistently negatively correlated with precipitation frequency in the observations and the twelve reanalysis products (Fig. 4). Based on the observations, the interannual variations in $T_a$ are jointly determined in part by precipitation frequency and $L_d$ in Northeast China and the northern part of Northwest China (Fig. 4). All of the reanalyses roughly capture these factors over these regions, although they display differences in the relative magnitudes (Fig. 4). Specifically, ERA-20CM, NOAA 20CRv2c, NOAA 20CRv2 and CFSR exhibit comparable relationships of $T_a$ with precipitation frequency and $L_d$; however, MERRA, MERRA2, NCEP-R2, ERA-20C,
and CERA-20C overestimate the relationship between $T_a$ and precipitation frequency, and ERA-Interim, JRA-55, and NCEP-R1 overestimate the relationship of $T_a$ with $L_d$ over these regions (Fig. 4).

Over the North China Plain and Middle China, the interannual variations in $T_a$ are partly jointly determined by $R_s$, precipitation frequency and $L_d$ (Fig. 4). The reanalyses roughly capture the effects of these three factors on $T_a$, although they display diverse combinations (Fig. 4). Among these combinations, JRA-55, MERRA2, ERA-20CM and ERA-Interim are comparable to the observations over these regions (Fig. 4). Over Southeast China, the interannual variations in $T_a$ are primarily regulated by $L_d$, precipitation frequency and $R_s$ (Fig. 4). The reanalyses exhibit slightly overestimated relationships of $T_a$ with $R_s$ and underestimated relationships with precipitation frequency (Fig. 4).

Over the Tibetan Plateau, the interannual variations in $T_a$ are regulated by $R_s$ and precipitation frequency (Fig. 4). Most of the reanalyses roughly capture the combinations of these factors but exhibit certain differences in the relative effects of $R_s$ and precipitation frequency on $T_a$ (Fig. 4). MERRA, MERRA2, NOAA 20CRv2c and NOAA 20CRv2 overestimate the relationships of $T_a$ with $R_s$ over the Tibetan Plateau (Fig. 4).

Overall, the spatial patterns of the simulated partial correlation of $T_a$ with $R_s$ in the reanalysis products are significantly correlated with those seen in the observations; $r=0.13-0.35$ ($p<0.05$) for the NWP-like reanalyses, and larger values of $r=0.24-0.41$ ($p<0.05$) are obtained for the climate reanalyses. Moreover, the spatial patterns in the
sensitivity of $T_a$ to $R_s$ exhibit significant correlations ($r=0.12-0.17$, $p<0.05$) for most of the climate reanalyses (Table 1). Precipitation frequency displays the largest spatial correlations ($r=0.16-0.43$, $p<0.05$) of the sensitivity of $T_a$ with these three relevant parameters in the reanalyses (Table 3). Significant spatial correlations reflecting the relationship (including the partial correlation and sensitivity) of $T_a$ with $L_d$ are also found (Table 1).

### 3.4 Regional Warming Trend Biases and Their Causes

#### 1) The Whole of China

From 1979 to 2010 over China, $T_a$ exhibits strong warming trends of 0.37°C/decade ($p<0.05$) in the observations and 0.22-0.48°C/decade ($p<0.05$) in the twelve reanalyses (Figs. 5 and S2-S3, Table 2). ERA-Interim and JRA-55 display spatial correlations with the observations ($r=0.47$ and 0.54, $p<0.05$) that are due at least partly to the inclusion of some $T_a$ observations, whereas NCEP-R2 and ERA-20C display the worst performance (Figs. S3, Tables 1 and 3). Furthermore, approximately 87% of the observed trends in $T_a$ over China can be explained by the greenhouse effect (i.e., 65% can be explained by the trend in $L_d$), precipitation frequency (29%) and $R_s$ (-7%, due to the trend in radiative forcing of -1.1 W·m⁻²/decade) (Figs. S3-4). The influence of the greenhouse effect on the observed trends in $T_a$ consists mainly of the trends in the atmospheric water vapor (42%) and the cloud fraction (3%) (Fig. S5). Among the reanalyses, over 90% of the trend in $T_a$ can be explained by the greenhouse effect, precipitation frequency and $R_s$ (Figs. S4-6). Specifically, ERA-Interim, JRA-55, MERRA and MERRA2 display the best ability to
capture the contributions of the greenhouse effect (48% to 76%), precipitation frequency (22% to 34%) and $R_s$ (-4% to 13%) to the trend in $T_a$ over China (Figs. S4 and S6). The remaining NWP-like reanalyses (i.e., NCEP-R1 and NCEP-R2) substantially overestimate the contribution of $R_s$ to the trend in $T_a$, whereas the climate reanalyses overestimate the contribution from $L_d$ (Figs. S4 and S6).

Here, we further quantify the contributions to the biases in the trend in $T_a$ made by those in $R_s$, $L_d$ and precipitation frequency among the twelve reanalyses over China and its seven subregions (Figs. 6-7). Over China, the overestimated $R_s$ trends (by 0.00-3.93 W·m$^{-2}$/decade; Figs. S8 and S13) increase the trends in $T_a$ (by 0.02-0.16°C/decade; Fig. 7) in the twelve reanalyses; the underestimated $L_d$ trends (by -0.25 to -1.61 W·m$^{-2}$/decade for the NWP-like reanalyses; Figs. S10 and S15) decrease the trends in $T_a$ (by -0.05 to -0.25°C/decade for the NWP-like reanalyses; Fig. 7); and the biases in the trends in precipitation frequency (by approximately -1.5 days/decade for the NWP-like reanalyses and approximately 2.6 days/decade for the climate reanalyses; Figs. S9 and S14) decrease the trends in $T_a$ (by 0.01 to 0.05°C/decade for the NWP-like reanalyses and -0.01 to -0.06°C/decade for the climate reanalyses; Fig. 7). Together, these effects produce an underestimate in the trends in $T_a$ on the order of 0.10°C/decade in the reanalyses (Fig. 7 and Table 2).

2) Seven Subregions

However, averaged trends over large areas may mask regional differences that reflect diverse regional warming biases and their causes (Figs. 5-7). The mean-adjusted spatial patterns of the biases in the trends in $T_a$ appear to be consistent
among the twelve reanalyses (Fig. S7) and mimic the spatial patterns in the overestimated $R_s$ trends over the North China Plain, South China and Northeast China (Fig. S8), given the spatial correlations between these variables in most of the reanalyses ($r=0.11-0.42$, $p<0.05$) (Figs. 6 and S7-8, Table 3). However, the reanalyses still underestimate the trends in $T_a$ over most of the regions. The key reason for this underestimation is the increase in precipitation frequency over Northwest China, the Loess Plateau, and Middle China seen in the NWP-like reanalyses and that seen over broader regions in the climate reanalyses (Figs. 5-6 and S9). This relationship is reflected by their negative spatial correlation, which has a maximum value of -0.62 ($p<0.05$) for MERRA (Table 3). Moreover, the decrease in $L_d$, which occurs due to the decreases in the atmospheric water vapor and cloud fraction that occur in the NWP-like reanalyses (Figs. S10-12), substantially cancels the warming effect of the overestimation of $R_s$ on $T_a$ over eastern China (Figs. 5 and S7). The opposite changes occur over Southeastern China in the climate reanalyses (Figs. 5 and S10). The effect of the changes in $L_d$ is reflected by its spatial correlations of up to 0.50 ($p<0.05$) (Table 3).

The corresponding contributions to the biases in the $T_a$ trend from are calculated from those in $R_s$, $L_d$ and precipitation frequency over seven subregions of China (Figs. 6-7).

Here, we further quantify the contributions to the biases in the trend in $T_a$ made by those in $R_s$, $L_d$ and precipitation frequency among the twelve reanalyses over China and its seven subregions (Figs. 6-7). Over China, the overestimated $R_s$ trends (by...
increase the trends in \( T_a \) (by 0.02-0.16°C/decade; Fig. 7) in the twelve reanalyses; the underestimated \( L_d \) trends (by -0.25 to -1.61 W·m\(^{-2}\)/decade for the NWP-like reanalyses; Figs. S10 and S15) decrease the trends in \( T_a \) (by 0.05 to 0.25°C/decade for the NWP-like reanalyses; Fig. 7); and the biases in the trends in precipitation frequency (by approximately 1.5 days/decade for the NWP-like reanalyses and approximately 2.6 days/decade for the climate reanalyses; Figs. S9 and S14) decrease the trends in \( T_a \) (by 0.01 to 0.05°C/decade for the NWP-like reanalyses and 0.01 to 0.06°C/decade for the climate reanalyses; Fig. 7). Together, these effects produce an underestimate in the trends in \( T_a \) on the order of 0.10°C/decade in the reanalyses (Fig. 7 and Table 2).

Over northern China, biases in the trend in \( T_a \) result primarily from those in precipitation frequency and \( L_d \) (Figs. 6-7). Over Northeast China, the observations exhibit an amplified warming of 0.41°C/decade (\( p<0.05 \); Fig. 4 and Table 2). This warming is significantly underestimated by NCEP-R1, JRA-55, NOAA 20CRv2 and NOAA 20CRv2c (by on the order of -0.15°C/decade) and is overestimated by MERRA and CFSR (by on the order of 0.2°C/decade) (Figs. 6-7). These biases in the trends in \( T_a \) in the reanalysis are jointly explained by the warming (0.04-0.48°C/decade) induced by the underestimated trends in precipitation frequency and the cooling (-0.04 to -0.42°C/decade) induced by the underestimated trends in \( L_d \) (Fig. 7).

Over Northwest China, the biases in the trend in precipitation frequency and \( L_d \) are mainly explained by the overestimated warming in NCEP-R2 (by 0.22°C/decade).
The substantially underestimated trend in $L_d$ induced by the decrease in the atmospheric water vapour and cloud fraction (Figs. S9-S12 and S16-17) lead to an underestimate of the warming in MERRA (by -0.22°C/decade) (Fig. 7).

Most of the reanalyses display weakened warming over the Tibetan Plateau and the Loess Plateau (Fig. 5 and S3, Table 2). In particular, NCEP-R1 and NCEP-R2 fail to reproduce the warming over the Tibetan Plateau, and MERRA fails to reproduce the warming over the Loess Plateau (Fig. 5 and S3, Table 2). The significant cooling biases in the trends in $T_a$ (by -0.02 to -0.31°C/decade) over the Tibetan Plateau and the Loess Plateau result from the underestimated trends in $L_d$ and the overestimated trends in precipitation frequency seen in most of the reanalyses (Figs. 5-7 and S9-12). These cooling biases are further induced by the underestimated trends in $R_s$ (Figs. 5-7 and S8).

Over southern China, the biases in the trend in $T_a$ are regulated by the biases in the trends in $R_s$, $L_d$ and precipitation frequency (Figs. 6-7). Over Southeast China, the significantly overestimated trends in $T_a$ (by 0.04, 0.02 and 0.17°C/decade, respectively) are induced by the overestimated trends in $R_s$ (by 4.25, 3.34 and 6.27 W·m$^{-2}$/decade, respectively) seen in ERA-Interim, JRA-55 and CFSR (Figs. 6-7 and S8). The underestimated trends in $T_a$ are induced by the overestimated trends in precipitation frequency and $L_d$ in NCEP-R1, MERRA, ERA-20CM, CERA-20C, NOAA 20CRv2 and NOAA 20CRv2c (Figs. 6-7 and S9).

Over Middle China, the significantly overestimated trends in $T_a$ (by 0.04, 0.06, 0.11, 0.03, 0.11 and 0.14°C/decade, respectively) are induced by the overestimated
trends in $R_s$ (by 2.09, 1.50, 2.59, 1.20 and 4.81 W·m$^{-2}$/decade, respectively) seen in ERA-Interim, JRA-55, ERA-20C, ERA-20CM, CERA-20C and CFSR (Figs. 6-7 and S8). The overestimated trends in precipitation frequency may lead to cooling in the trends in $T_a$ in the reanalyses, especially for MERRA (which reflects an induced bias in the trend of -0.15°C/decade) over Middle China (Figs. 6-7 and S9).

Due to the underestimated trends in the atmospheric water vapor and the cloud fraction (Figs. S11-12), the underestimation of $L_d$ produces a cooling effect on the trend in $T_a$ (by -0.05 to -0.32°C/decade) in the reanalyses over the North China Plain (Figs. 6-7 and S10). However, due to the lack of inclusion of plausible trends in aerosol loading, the substantial increases in $R_s$ over the North China Plain (Fig. S8) have strong warming effects on the trends in $T_a$ (by 0.01 to 0.21°C/decade) in the reanalyses (Figs. 6-7 and S8). The biases in the trends in precipitation frequency (of approximately -2.5 days/decade for the NWP-like reanalyses and approximately 1.5 days/decade for some of the climate reanalyses) contribute some part of the biases in the trends in $T_a$ (approximately 0.05°C/decade for the NWP-like reanalyses and -0.03°C/decade for the climate reanalyses).

Overall, the biases in the trends in $T_a$ in the reanalyses can be substantially explained by those in $L_d$, precipitation frequency and $R_s$, but this effect varies regionally (Figs. 6-7). Over northern China, the biases in the trend in $T_a$ (which are on the order of -0.12°C/decade) result primarily from a combination of those in $L_d$ (which are on the order of -0.10°C/decade) and precipitation frequency (which are on the order of 0.05°C/decade), with relatively small contributions from $R_s$ (which are on
the order of -0.03°C/decade. Over southern China, the biases in the trend in $T_a$ (which are on the order of -0.07°C/decade) are caused by those in $R_s$ (which are on the order of 0.10°C/decade), $L_d$ (which are on the order of -0.08°C/decade) and precipitation frequency (which are on the order of -0.06°C/decade) (Fig. S18). Note also that the incorporation of the observed changes in surface air temperatures in ERA-Interim and JRA-55 may introduce biases into the trends in the output $T_a$ values; however, the use of partial correlation and regression analysis would lead to smaller impacts of the biases in these physical variables in quantifying their contributions to the trends in $T_a$.

3.5 Spatial Linkages of Biases in the Warming Trends in the Twelve Reanalyses

We next integrate the relationships of the spatial patterns in the biases in the trends in $T_a$ with those in $R_s$, $L_d$ and precipitation frequency over China in the twelve reanalyses (Fig. 8). The results show that the biases in the trends in $T_a$ show significant correlations with $R_s$ ($r=0.80$, slope=0.06, $p=0.09$) and precipitation frequency ($r=-0.83$, slope=-0.04, $p=0.02$) and $L_d$ ($r=0.77$, slope=0.10, $p=0.10$) in the twelve reanalyses if information on these patterns is included. When the spatial patterns of the biases in the trends in these variables are not considered, the biases in the trends in $T_a$ show relatively small correlations with $R_s$ ($r=0.32$, slope=0.02, $p>0.1$), precipitation frequency ($r=-0.51$, slope=-0.02, $p=0.09$) and $L_d$ ($r=0.14$, slope=0.02, $p>0.1$) in the reanalyses (Fig. 8). Similar results are obtained for the atmospheric water vapor ($r=0.71$, $p=0.1$) and the cloud fraction ($r=-0.74$, $p=0.09$) if their spatial patterns are considered (Figs. S19), and this relationship involving the cloud fraction
is very similar to that associated with $R_s$ (Figs. 8 and S19). Within the subregions of China, the biases in the trends in $T_a$ show significant correlations with $R_s$ ($r=0.68$ to 0.90, $p<0.1$), precipitation frequency ($r=-0.55$ to -0.94, $p<0.1$) and $L_d$ ($r=0.53$ to 0.93, $p<0.1$) when the spatial patterns in the reanalyses are included (Fig. S20). These results provide a novel perspective that can be used to investigate the spatial relationships between biases in the trends in $T_a$ and relevant quantities in reanalyses.

4. Discussion

In this section, we first examine the possible impacts of data homogenization on the trends in $T_a$. The trends in $T_a$ derived from the original dataset are almost as high as those from the homogenized dataset, especially over the North China Plain and Northwest China (Fig. 5 and Table 2). Homogenization primarily adjusts breakpoints in time series (Wang, 2008), which occur mainly due to station relocation and changes in instruments (Cao et al., 2016; Li et al., 2017; Wang, 2014), and it helps to objectively depict trends in $T_a$, thus permitting the assessment of the modelled trends in $T_a$ and its spatial patterns that are present in the reanalyses.

We found that the elevation differences between the models and the stations influence the biases in the trends in $T_a$ but cannot explain the spatial patterns in the biases in the trends in $T_a$ (average $r=0.11$) (Fig. S21). Comparison of the models that use the same grid (NOAA 20CRv2c vs. NOAA 20CRv2, MERRA vs. MERRA2, NCEP-R1 vs. NCEP-R2 and ERA-20C vs. ERA-20CM) shows that the one is correlated with elevation differences, but the other is not, which implies that this
statistical correlation does not have physical significance. In addition, elevation differences do not change with time. Nevertheless, the spatial patterns in the normalized trends in $T_a$ (excluding the impacts of the absolute value of temperature on the trends) are very near to those of the trends (Fig. S22), implying that the differences in the absolute value of temperature have an important effect, given that the site-to-grid inconsistency can be neglected.

In the reanalyses, vegetation is only included as climatological information, but the vegetation displays a growth trend during the study period of 1979-2010 within China (Fig. S23). This discrepancy positively enlarges the biases in the trends in $T_a$ due to the vegetation cooling effect (Zeng et al., 2017; Trigo et al., 2015). This effect is reflected by the negative spatial correlation ($r=-0.26$, $p=0.00$) between the inverted trend in the NDVI and the biases in the trend in $T_a$ (Fig. S23). The growth of vegetation reduces $T_a$ by regulating surface roughness, surface conductivity, soil moisture and albedo to partition greater amounts of available energy into latent heat fluxes, which leads to the formation of more precipitation (Shen et al., 2015; Spracklen et al., 2013). Thus, the inclusion of vegetation growth will improve the simulation of trends and especially the spatial pattern of $T_a$ in the reanalyses through the incorporation of more complete physical parameterizations (Li et al., 2005; Dee and Todling, 2000; Trigo et al., 2015).

Due to their inclusion of surface air temperature observations, ERA-Interim and JRA-55 display high skill in reproducing the observed patterns; they have near-zero means (0.01 and 0.01°C/decade) and the smallest standard deviations (0.16 and
0.15°C/decade) of the trend biases among the twelve reanalysis products. However, pattern differences of 37.8% (standard deviation of trend bias/China-averaged trend) are still evident (Figs. 5 and 8). Although it does not incorporate surface air temperature observations, ERA-20CM presents a pattern (with a mean of -0.04°C/decade and a standard deviation of 0.15°C/decade; Figs. 5 and 8) that is comparable to those of ERA-Interim and JRA-55 and better than that of ERA-20C (mean of -0.08°C/decade and standard deviation of 0.20°C/decade; Figs. 5 and 8), which uses the same forecast model as ERA-20CM. These results imply that ensemble forecasting could be used to meet important goals. The ensemble forecasting simulation technique used in ERA-20CM also displays advantages in that it yields an improved simulated pattern of biases in the trends in $R_s$ (SD=1.84 W·m$^{-2}$/decade, 171%), precipitation frequency (SD=2.78 days/decade, 122%) and $L_d$ (SD=1.25 W·m$^{-2}$/decade, 82%) (Fig. 8).

We consider the degree to which the ensemble assimilation technique can improve the spatial patterns of the biases in the trends in $T_a$ in the reanalyses. We find that this technique can detect the biases in the trends in $T_a$ over more another approximately 12% (8%) of the grid cells in CERA-20C, which incorporates 10 ensemble members (NOAA 20CR2vc and NOAA 20CR2v employ 56 ensemble members) (Figs. 5 l-n). However, the biases in the trends in $T_a$ over these grid cells are not significant at a significance level of 0.05, according to Student’s $t$-test, implying that the ensemble assimilation technique cannot explain the spatial pattern of the biases in the trends in $T_a$ identified in this study (in Figs. 5 l-n).
To provide a preliminary discussion of the improvements in climate forecast models in reflecting patterns in climate trends, we compare the spatial patterns of the biases in the trends in $R_s$, precipitation frequency and $L_d$ in the reanalyses because observations of these variables are not incorporated in the reanalyses. We find that the climate forecast models, i.e., ERA-20C, ERA-20CM, CERA-20C, NOAA 20CRv2c and NOAA 20CRv2, display better performance in reproducing the pattern of biases in the trends in $R_s$ (mean of 1.36 vs. 2.18 W·m$^{-2}$/decade; SD of 2.04 vs. 2.71 W·m$^{-2}$/decade), precipitation frequency (mean of 1.32 vs. -1.44%/decade; SD of 3.57 vs. 6.14%/decade) and $L_d$ (mean of 0.12 vs. -0.85 W·m$^{-2}$/decade; SD of 1.33 vs. 1.50 W·m$^{-2}$/decade) than the NWP-like models, i.e., ERA-Interim, NCEP-R1, MERRA, JRA-55, NCEP-R2 and MERRA2 (Fig. 8). In addition, because the SST boundary condition evolves freely in CFSR, the patterns of biases in the trends in $R_s$, precipitation frequency and $L_d$ in CFSR differ substantially from those in the other reanalyses.

We also consider whether the spatial pattern of biases in the trend in $T_a$ is altered by the atmospheric circulation patterns simulated by the ERA-20CM ensemble. In ERA-20CM, the atmospheric circulation patterns are influenced by SSTs and sea ice and then partly mediate the influence of global forcings on the trends in $T_a$. In ERA-20CM, the probability distribution function of the biases in the trends in $T_a$ from outside the ensemble ranges incorporates that from Student’s $t$-test at a significance level of 0.05 (Fig. 5k). This result has important implications in that 1) the climate variability in the ensembles under the different model realizations of SSTs and sea ice
cover does not change the pattern of the biases in the trends in $T_a$ (Fig. 5k); moreover,

2) Student’s $t$-test exhibits a suitable ability to detect the significance of the biases in
the trends in $T_a$ (Fig. 5k) when considering the effects of interannual variability on the
trend.

Overall, producing global or regional reanalyses that adequately reflect regional
climate is challenging using the current strategy, and further improvements are
required. The results and discussion above indicate some potential but challenging
approaches that can be used to maximize the signal component corresponding to the
regional climate in final reanalyses and robustly narrow the uncertainties in trends.

1) MERRA2’s pioneering incorporation of time-varying aerosol loadings
provides a way of improving the representation of regional temperature changes over
regions such as the North China Plain where the impacts of aerosols on surface
temperatures are significant. MERRA2 incorporates time-varying aerosol loadings in a
pioneering attempt to improve regional warming over the North China Plain to some
extent. Thus, we encourage research groups to include accurate aerosol information
and improve the skill of simulation of the energy budget and partitioning, especially
of regional surface incident solar radiation, in other reanalyses.

2) To improve regional climate modelling, forecast output should be produced
using a physical ensemble like that employed in ERA-20CM to quantify the
uncertainties associated with the relevant parameterizations in the reanalyses, due to
the impossibility of optimizing all of the biases. Meanwhile, careful ensemble design
would likely yield useful information for use in improving models, assimilation
methods and the bias correction of observations by exploring the interdependency among sources of errors. Such designs would undoubtedly have additional benefits for further development, leading to the next generation of reanalyses.

3) To improve coupled land-atmospheric interactions, the true dynamics of land cover and use should be incorporated. Moreover, the physical parameterizations should be improved, including the responses of surface roughness, surface conductivity and albedo to regional climate. These changes would represent an improvement over the use of constant types and fractions of vegetation, as is done in ERA-Interim (Zhou and Wang, 2016b).

4) Given the implications of the spurious performance of the freely evolving boundary conditions in CFSR, homogeneous and accurate records of SST and sea ice should be produced.

Next-generation reanalyses, including both global and regional reanalyses, will assimilate and analyse in situ observations, satellite radiance, and other remote observations. In addition to short-term accuracy and long-term trends, they will also focus on spatial patterns by incorporating or improving accurate representations of land surface conditions and processes within the coupled weather and climate Earth systems. Thus, these reanalyses will advance the simulation of land-atmosphere interactions to yield high skill in studies of regional warming and the detection and attribution of regional climate change using various datasets, which frequently include global and regional reanalyses (Zhou et al., 2018; Zhou and Wang, 2016d; Herring et al., 2018; Trenberth et al., 2015; Stott, 2016; Dai et al., 2017; Zhou and Wang, 2017b).
Additionally, the uncertainties associated with regional warming could be ascertained using physics ensembles with various equiprobable realizations of boundary conditions.

5. Conclusions

The reanalyses display differences in $T_a$ when compared to the observations with a range of -10–10°C over China. Approximately 74% and 6% of these differences can be explained by site-to-grid elevation differences and the filtering error in the elevations used in the spectral models. These results imply fairly good skill in the simulation of the climatology of $T_a$ in the twelve reanalyses over China. Moreover, the twelve reanalyses roughly capture the interannual variability in $T_a$ (median $r=0.95$). In the reanalyses, $T_a$ displays a consistently positive correlation with $R_s$ and $L_d$ and is negatively correlated with precipitation frequency, as seen in observations, despite the evident spatial patterns in their magnitudes over China.

$T_a$ exhibits a strong warming trend of 0.37°C/decade ($p<0.05$) in the observations and 0.22-0.48°C/decade ($p<0.05$) in the twelve reanalyses over China. In the observations, approximately 87% of the observed trend in $T_a$ over China can be explained by the greenhouse effect (i.e., 65% can be explained by the trend in $L_d$), precipitation frequency (29%) and $R_s$ (-7%, due to the trend in radiative forcing of -1.1 W·m$^{-2}$/decade).

However, the biases in the trends in $T_a$ seen in the reanalyses relative to the observations display an evident spatial pattern (mean=-0.16~0.11°C/decade,
The spatial patterns of the biases in the trends in the values of $T_a$ in the reanalyses are significantly correlated with those in $R_s$ (maximum $r=0.42$, $p<0.05$), precipitation frequency (maximum $r=-0.62$, $p<0.05$) and $L_d$ (maximum $r=0.50$, $p<0.05$). Over northern China, the biases in the trends in $T_a$ (which are on the order of -0.12°C/decade) result primarily from a combination of those in $L_d$ (which are on the order of -0.10°C/decade) and precipitation frequency (which are on the order of 0.05°C/decade), with relatively small contributions from $R_s$ (which are on the order of -0.03°C/decade). Over southern China, the biases in the trends in $T_a$ (which are on the order of -0.07°C/decade) are regulated by the biases in the trends in $R_s$ (which are on the order of 0.10°C/decade), $L_d$ (which are on the order of -0.08°C/decade) and precipitation frequency (which are on the order of -0.06°C/decade).

If information on spatial patterns is included, the simulated biases in the trends in $T_a$ correlate well with those of precipitation frequency, $R_s$ and $L_d$ in the reanalyses ($r=-0.83$, 0.80 and 0.77, $p<0.1$); similar results are obtained for the atmospheric water vapor and the cloud fraction ($r=0.71$ and -0.74, $p<0.1$). These results provide a novel perspective that can be used to investigate the spatial relationships between the biases in the trends in $T_a$ and the relevant parameters among the twelve reanalyses. Therefore, improving simulations of precipitation frequency and $R_s$ helps to maximize the signal component corresponding to the regional climate. In addition, the analysis of $T_a$ observations helps to improve the performance of regional warming in ERA-Interim and JRA-55. Incorporating vegetation dynamics in reanalyses and the use of accurate...
aerosol information, as in MERRA-2, would advance the modelling of regional warming. The ensemble technique adopted in ERA-20CM, a twentieth-century atmospheric model ensemble that does not assimilate observations, significantly narrows the uncertainties of regional warming in the reanalyses (standard deviation=0.15°C/decade).

Acknowledgements This study was funded by the National Key R&D Program of China (2017YFA0603601) and the National Natural Science Foundation of China (41525018). The latest observational datasets were obtained from the China Meteorological Administration (CMA; http://www.cma.gov.cn). Considerable gratitude is owed to several reanalysis working teams, including the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ERA-Interim, ERA-20C, ERA-20CM and CERA-20C data (http://www.ecmwf.int/); the Global Modelling and Assimilation Office (GMAO) at the NASA Goddard Space Flight Center for providing the MERRA and MERRA2 data (http://gmao.gsfc.nasa.gov/); the NOAA Earth System Research Laboratory (ESRL) for providing the NCEP-R1, NCEP-R2, CFSR, NOAA 20CRv2 and NOAA 20CRv2c data (http://www.esrl.noaa.gov/); and the Climate Prediction Division of the Global Environment and Marine Department at the Japan Meteorological Agency for providing the JRA-55 data (http://jra.kishou.go.jp/). We thank the Expert Team on Climate Change Detection and Indices (ETCCDI) for providing the RHtestV4 package (http://etccdi.pacificclimate.org/software.shtml), the United States
Geological Survey Earth Resources Observation and Science Data Center for providing the GTOPO30 data (http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html) and the working team of the Global Inventory Monitoring and Modelling System (GIMMS) project (https://ecocast.arc.nasa.gov/data/pub/gimms/). We thank Kevin E. Trenberth for his insightful suggestions.
References


1998.


Mao, J., Shi, X., Ma, L., Kaiser, D. P., Li, Q., and Thornton, P. E.: Assessment of reanalysis daily extreme temperatures with china's homogenized historical dataset


Zhou, C., and Wang, K.: Spatiotemporal divergence of the warming hiatus over land based on different definitions of mean temperature, Sci. Rep., 6, 31789, 10.1038/srep31789, 2016d.


Table 1. Summary information on the twelve reanalysis products, including institution, model resolution, assimilation system, surface observations included associated with surface air temperatures, sea ice and sea surface temperatures (SSTs) and greenhouse gas (GHG) boundary conditions. The number in the parentheses in the Model Name column is the year of the version of the forecast model used. More details on each product can be found in the associated reference.

<table>
<thead>
<tr>
<th>Reanalysis Institution</th>
<th>Model Name</th>
<th>Model Resolution</th>
<th>Period</th>
<th>Assimilation System</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERA-Interim ECMWF</td>
<td>IFS version Cy31r2 (2007)</td>
<td>T255 – 80 km, 60 levels</td>
<td>1979 onwards</td>
<td>4D-VAR</td>
</tr>
<tr>
<td>JRA-55 JMA</td>
<td>JMA operational numerical weather prediction system (2009)</td>
<td>T319 – 55 km, 60 levels</td>
<td>1958-2013</td>
<td>4D-VAR</td>
</tr>
<tr>
<td>NCEP-R1 NCEP/NCAR</td>
<td>NCEP operational numerical weather prediction system (1995)</td>
<td>T62 – 210 km, 28 levels</td>
<td>1948 onwards</td>
<td>3D-VAR</td>
</tr>
<tr>
<td>NCEP-R2 NCEP/DOE</td>
<td>Modified NCEP-R1 model (1998)</td>
<td>T62 – 210 km, 28 levels</td>
<td>1979 onwards</td>
<td>3D-VAR</td>
</tr>
<tr>
<td>MERRA NASA/GMAO</td>
<td>GEOS-5.0.2 atmospheric general circulation model (2008)</td>
<td>0.5°× 0.667°, ~55 km, 72 levels</td>
<td>1979 onwards</td>
<td>3D-VAR with incremental updating (GEOS IAU)</td>
</tr>
<tr>
<td>MERRA-2 NASA/GMAO</td>
<td>Updated version of GEOS-5.12.4 used in MERRA; its land model is similar to that of MERRA (2015)</td>
<td>0.5°× 0.625°, ~55 km, 72 levels</td>
<td>1980 onwards</td>
<td>3D-VAR with incremental updating (GEOS IAU)</td>
</tr>
<tr>
<td>ERA-20C ECMWF</td>
<td>IFS version Cy38r1 (2012), coupled atmosphere-land-ocean-waves system</td>
<td>T159 – 125 km, 91 levels</td>
<td>1900-2010</td>
<td>4D-VAR</td>
</tr>
<tr>
<td>ERA-20CM ECMWF</td>
<td>Similar to that used in ERA-20C (2012)</td>
<td>T159 – 125 km, 91 levels</td>
<td>1900-2010</td>
<td>4D-VAR</td>
</tr>
<tr>
<td>CERA-20C ECMWF</td>
<td>IFS version Cy41r2 (2016), coupled atmosphere-ocean-land-waves-sea ice system</td>
<td>T159 – 125 km, 91 levels</td>
<td>1901-2010</td>
<td>CERA ensemble assimilation technique</td>
</tr>
<tr>
<td>NOAA 20CRv2c NOAA/ESRL PSD</td>
<td>NCEP GFS (2008), an updated version of the NCEP Climate Forecast System (CFS) coupled atmosphere-land model</td>
<td>T62 – 210 km, 28 levels</td>
<td>1851-2014</td>
<td>Ensemble Kalman filter</td>
</tr>
<tr>
<td>CFSR NCEP</td>
<td>NCEP CFS (2011) coupled atmosphere-ocean-land-sea ice model</td>
<td>T382 – 38 km, 64 levels</td>
<td>1979-2010</td>
<td>3D-VAR</td>
</tr>
<tr>
<td>Related Assimilated and Analysed Observations</td>
<td>Sea Ice and SSTs</td>
<td>GHG Forcing</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1) Includes <em>in situ</em> observations of near-surface air temperature/pressure/relative humidity</td>
<td>A changing suite of SST and sea ice data from observations and NCEP</td>
<td>Interpolation by 1.6 ppmv/year from the global mean CO₂ in 1990 of 353 ppmv</td>
<td>(Dee et al., 2011b)</td>
<td></td>
</tr>
<tr>
<td>2) Assimilates upper-air temperatures/wind/specific humidity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Assimilates rain-affected SSMI radiances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Analyses available near-surface observations</td>
<td><em>In situ</em> observation-based estimates of the COBE SST data and sea ice</td>
<td>Same as CMIP5</td>
<td>(Kobayashi et al., 2015)</td>
<td></td>
</tr>
<tr>
<td>2) Assimilates all available traditional and satellite observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Initiated with weather observations from ships, planes, station data, satellite observations and many more sources</td>
<td>Reynolds SSTs for 1982 on and the UKMO GISST data for earlier periods; sea ice from SMMR/SSMI</td>
<td>Constant global mean CO₂ of 330 ppmv; no other trace gases</td>
<td>(Kalnay et al., 1996)</td>
<td></td>
</tr>
<tr>
<td>2) No inclusion of near-surface air temperatures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Uses observed precipitation to nudge soil moisture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) No information on aerosols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) No inclusion of near-surface air temperatures</td>
<td>AMIP-II prescribed</td>
<td>Constant global mean CO₂, 350 ppmv; no other trace gases</td>
<td>(Kanamitsu et al., 2002)</td>
<td></td>
</tr>
<tr>
<td>2) No information on aerosols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Neither MERRA nor MERRA-2 analyse near-surface air temperature, relative humidity, or other variables</td>
<td>Assimilates data and includes radiative forcings from CMIP5</td>
<td>Same as CMIP5</td>
<td>(Rienecker et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>2) Radiosondes do provide some low-level observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Includes newer observations (not included in MERRA) after the 2010s</td>
<td>AMIP-II and Reynolds SSTs</td>
<td>Same as CMIP5</td>
<td>(Reichle et al., 2017)</td>
<td></td>
</tr>
<tr>
<td>2) Includes aerosols from MODIS and AERONET measurements over land after the 2000s and from the GOCART model before the 2000s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Assimilates observation-corrected precipitation to correct the model-generated precipitation before reaching the land surface</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Assimilates surface pressures from ISPDv3.2.6 and ICOADSv2.5.1 and surface marine winds from ICOADSv2.5.1</td>
<td>SSTs and sea ice from HadISST2.1.0.0</td>
<td>Same as CMIP5</td>
<td>(Poli et al., 2016)</td>
<td></td>
</tr>
<tr>
<td>2) Uses monthly climatology of aerosols from CMIP5</td>
<td>SSTs and sea ice realizations from HadISST2.1.0.0 used in 10 members</td>
<td>Same as CMIP5</td>
<td>(Hersbach et al., 2015)</td>
<td></td>
</tr>
<tr>
<td>Assimilates no data and includes radiative forcings from CMIP5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Assimilates surface pressures from ISPDv3.2.6 and ICOADSv2.5.1 and surface marine winds from ICOADSv2.5.1</td>
<td>SSTs from HadISST2.1.0.0</td>
<td>Same as CMIP5</td>
<td>(Laloyaux et al., 2016)</td>
<td></td>
</tr>
<tr>
<td>2) Assimilates no data in the land, wave and sea ice components but uses the coupled model at each time step</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assimilates only surface pressure and sea level pressure</td>
<td>SSTs from HadISST1.1 and sea ice from COBE SST</td>
<td>Monthly 15° gridded estimates of CO₂ from WMO observations</td>
<td>(Compo et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>Same as NOAA 20CRv2c</td>
<td>SSTs and sea ice from HadISST1.1</td>
<td>Monthly 15° gridded estimates of CO₂ from WMO observations</td>
<td>(Compo et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>1) Assimilates all available conventional and satellite observations but not near-surface air temperatures</td>
<td>Generated by coupled ocean-sea ice models; evolves freely during the 6-h coupled model integration</td>
<td>Monthly 15° gridded estimates of CO₂ from WMO observations</td>
<td>(Saha et al., 2010)</td>
<td></td>
</tr>
<tr>
<td>2) Atmospheric model contains observed changes in aerosols</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Uses observation-corrected precipitation to force the land surface analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2.

Differences (unit: °C) relative to the homogenized observations and trends (unit: °C/decade) in surface air temperatures ($T_a$) from 1979 to 2010 over China and its seven subregions. The bold and italic bold fonts indicate results that are significant according to two-tailed Student’s $t$-tests with significance levels of 0.05 and 0.1, respectively.

<table>
<thead>
<tr>
<th>Region</th>
<th>China</th>
<th>Tibet Plateau</th>
<th>Northwest China</th>
<th>Loess Plateau</th>
<th>Middle China</th>
<th>Northeast China</th>
<th>North China Plain</th>
<th>Southeast China</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERA-Interim</td>
<td>-0.87</td>
<td>0.38</td>
<td>-3.49</td>
<td>0.33</td>
<td>-1.82</td>
<td>0.37</td>
<td>-0.32</td>
<td>0.50</td>
</tr>
<tr>
<td>NCEP-R1</td>
<td>-2.56</td>
<td>0.23</td>
<td>-6.80</td>
<td>0.11</td>
<td>-4.45</td>
<td>0.39</td>
<td>-1.77</td>
<td>0.21</td>
</tr>
<tr>
<td>MERRA</td>
<td>-0.48</td>
<td>0.25</td>
<td>-3.48</td>
<td>0.33</td>
<td>0.95</td>
<td>0.14</td>
<td>1.14</td>
<td>0.09</td>
</tr>
<tr>
<td>JRA-55</td>
<td>-1.10</td>
<td>0.38</td>
<td>-3.49</td>
<td>0.42</td>
<td>-1.70</td>
<td>0.39</td>
<td>-0.58</td>
<td>0.52</td>
</tr>
<tr>
<td>NCEP-R2</td>
<td>-2.10</td>
<td>0.25</td>
<td>-5.76</td>
<td>0.07</td>
<td>-4.29</td>
<td>0.58</td>
<td>-1.33</td>
<td>0.10</td>
</tr>
<tr>
<td>MERRA2</td>
<td>-0.91</td>
<td>0.28</td>
<td>-3.41</td>
<td>0.35</td>
<td>0.34</td>
<td>0.32</td>
<td>0.12</td>
<td>0.19</td>
</tr>
<tr>
<td>ERA-20C</td>
<td>-1.42</td>
<td>0.29</td>
<td>-6.56</td>
<td>0.33</td>
<td>-1.95</td>
<td>0.31</td>
<td>0.03</td>
<td>0.21</td>
</tr>
<tr>
<td>ERA-20CM</td>
<td>-1.48</td>
<td>0.32</td>
<td>-5.93</td>
<td>0.28</td>
<td>-1.39</td>
<td>0.38</td>
<td>-0.36</td>
<td>0.33</td>
</tr>
<tr>
<td>NOAA 20Cv2c</td>
<td>-2.06</td>
<td>0.34</td>
<td>-7.00</td>
<td>0.41</td>
<td>-2.15</td>
<td>0.38</td>
<td>-0.78</td>
<td>0.36</td>
</tr>
<tr>
<td>NOAA 20Cv2</td>
<td>-0.28</td>
<td>0.22</td>
<td>-2.75</td>
<td>0.39</td>
<td>-0.01</td>
<td>0.28</td>
<td>1.62</td>
<td>0.16</td>
</tr>
<tr>
<td>CFSR</td>
<td>-0.32</td>
<td>0.24</td>
<td>-2.78</td>
<td>0.33</td>
<td>-0.01</td>
<td>0.29</td>
<td>1.48</td>
<td>0.20</td>
</tr>
<tr>
<td>Obs-raw</td>
<td>-1.74</td>
<td>0.48</td>
<td>-5.09</td>
<td>0.46</td>
<td>-1.03</td>
<td>0.44</td>
<td>-0.25</td>
<td>0.40</td>
</tr>
<tr>
<td>Obs-homogenized</td>
<td>0.03</td>
<td>0.40</td>
<td>0.03</td>
<td>0.46</td>
<td>0.09</td>
<td>0.44</td>
<td>0.01</td>
<td>0.52</td>
</tr>
</tbody>
</table>
Table 3. Spatial pattern correlation (unit: 1) of three groups: partial relationships, trends and simulated biases in the trends in surface air temperature ($T_a$) against surface incident solar radiation ($R_s$), precipitation frequency (PF) and surface downward longwave radiation ($L_d$). The bold and italic bold fonts indicate results that are significant according to two-tailed Student’s t-tests with significance levels of 0.05 and 0.1, respectively.

<table>
<thead>
<tr>
<th>Pattern Correlation</th>
<th>Partial Relationship</th>
<th>Trend</th>
<th>Trend Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($T_a$, $R_s$)</td>
<td>($T_a$, PF)</td>
<td>($T_a$, $L_d$)</td>
</tr>
<tr>
<td>ERA-Interim</td>
<td>0.29</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>NCEP-R1</td>
<td>0.30</td>
<td>0.06</td>
<td>0.18</td>
</tr>
<tr>
<td>MERRA</td>
<td>0.29</td>
<td>0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>JRA-55</td>
<td>0.35</td>
<td>0.21</td>
<td>0.22</td>
</tr>
<tr>
<td>NCEP-R2</td>
<td>0.22</td>
<td>0.03</td>
<td>0.20</td>
</tr>
<tr>
<td>MERRA2</td>
<td>0.13</td>
<td>0.05</td>
<td>0.26</td>
</tr>
<tr>
<td>ERA-20C</td>
<td>0.28</td>
<td>-0.07</td>
<td>-0.07</td>
</tr>
<tr>
<td>ERA-20CM</td>
<td>0.24</td>
<td>-0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>CERA-20C</td>
<td>0.41</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td>NOAA 20CRv2</td>
<td>0.39</td>
<td>0.15</td>
<td>-0.22</td>
</tr>
<tr>
<td>NOAA 20CRv2</td>
<td>0.38</td>
<td>0.15</td>
<td>-0.21</td>
</tr>
<tr>
<td>CFSR</td>
<td>0.33</td>
<td>0.12</td>
<td>0.10</td>
</tr>
<tr>
<td>Obs-raw</td>
<td>-0.07</td>
<td>0.27</td>
<td>0.50</td>
</tr>
<tr>
<td>Obs-homogenized</td>
<td>-0.09</td>
<td>0.35</td>
<td>0.32</td>
</tr>
</tbody>
</table>
Figure Captions:

Figure 1. The multiyear-averaged differences in surface air temperatures ($T_a$, unit: °C) during the period of 1979-2010 from the twelve reanalysis products relative to the homogenized observations over China. The reanalysis products are (a) ERA-Interim, (b) NCEP-R1, (c) MERRA, (d) JRA-55, (e) NCEP-R2, (f) MERRA2, (g) ERA-20C, (h) ERA-20CM, (i) CERA-20C, (j) NOAA 20CRv2c, (k) NOAA 20CRv2 and (l) CFSR. The mainland of China is divided into seven regions (shown in Fig. 1c), specifically ① the Tibetan Plateau, ② Northwest China, ③ the Loess Plateau, ④ Middle China, ⑤ Northeast China, ⑥ the North China Plain and ⑦ South China.

Figure 2. The impact of inconsistencies between station and model elevations on the simulated multiyear-averaged differences in surface air temperatures ($T_a$, unit: °C) during the study period of 1979-2010 over China. The elevation difference ($\Delta$Height) between the stations and the models consists of the filtering error in the elevations used in the spectral models ($\Delta f$) and the difference in site-to-grid elevations ($\Delta s$) due to the complexity of orographic topography. $\Delta f$ is derived from the model elevations minus the ‘true’ elevations in the corresponding model grid cells from GTOPO30. The GTOPO30 orography is widely used in reanalyses, e.g., by ECMWF. The colour bar denotes the station elevations (unit: m). The relationship of the $T_a$ differences is regressed on $\Delta$Height (shown at the bottom of each subfigure) or $\Delta f$ and $\Delta s$ (shown at the top of each subfigure); the corresponding explained variances are shown.

Figure 3. Taylor diagrams for annual time series of the observed and reanalysed surface air temperature anomalies ($T_a$, unit: °C) from 1979 to 2010 in (a) China and
the seven subregions. The correlation coefficient, standard deviation and root mean squared error (RMSE) are calculated against the observed homogenized $T_a$ anomalies.

**Figure 4.** Composite map of partial correlation coefficients of the detrended surface air temperature ($T_a$, unit: °C) against surface incident solar radiation ($R_s$), precipitation frequency (PF) and surface downward longwave radiation ($L_d$) during the period of 1979-2010 from observations and the twelve reanalysis products. The marker ‘+’ denotes the negative partial correlations of $T_a$ with $R_s$ over the Tibetan Plateau in NCEP-R2, ERA-20C and ERA-20CM.

**Figure 5.** (a, b) The observed trends in surface air temperature ($T_a$, unit: °C/decade) and the simulated biases in the trends in $T_a$ (unit: °C/decade) during the period of 1979-2010 from (c) raw observations and (d-o) the twelve reanalysis products over China with respect to the homogenized observations. The squares denote the original homogeneous time series, and the dots denote the adjusted homogeneous time series. The probability distribution functions of all of the biases in the trends are shown as coloured histograms, and the black stairs are integrated from the trend biases with a significance level of 0.05 (based on two-tailed Student’s $t$-tests). The cyan and green stars in (k-n) represent estimates of the biases in the trends outside the ensemble ranges whose locations are denoted by the black dots shown in (k-n).

**Figure 6.** Composite map of the contributions (unit: °C/decade) of the biases in the trends in three relevant parameters, surface incident solar radiation ($R_s$, in red), surface downward longwave radiation ($L_d$, in green) and precipitation frequency (in...
blue) to the biases in the trends in surface air temperature ($T_a$) during the study period of 1979-2010, as estimated using the twelve reanalysis products over China.

**Figure 7.** Contribution (unit: °C/decade) of the biases in the trends in surface air temperatures ($T_a$) from three relevant parameters, surface incident solar radiation ($R_s$, in brown), surface downward longwave radiation ($L_d$, in light blue) and precipitation frequency (PF, in deep blue) during the study period of 1979-2010 from the twelve reanalysis products over China and its seven subregions.

**Figure 8.** Spatial associations of the simulated biases in the trend in surface air temperature ($T_a$) versus three relevant parameters among the twelve reanalysis products (solid lines indicate the NWP-like reanalyses, and dashed lines indicate the climate reanalyses). The probability density functions (unit: %) of these biases in the trends are estimated from approximately 700 1°×1° grid cells that cover China. The median values (coloured dots with error bars of spatial standard deviations) of the biases in the trends in $T_a$ (unit: °C/decade) in the twelve reanalyses are regressed onto those of (a) the surface incident solar radiation ($R_s$, unit: $W \cdot m^{-2}$/decade), (b) precipitation frequency (unit: days/decade) and (c) the surface downward longwave radiation ($L_d$, unit: $W \cdot m^{-2}$/decade) using the ordinary least squares method (OLS, denoted by the dashed grey lines) and the weighted total least squares method (WTLS, denoted by the solid black lines). The 5-95% confidence intervals of the regressed slopes obtained using WTLS are shown as shading. The regressed correlations and slopes are shown as grey and black text, respectively.
Figure 1. The multiyear-averaged differences in surface air temperatures ($T_a$, unit: °C) during the period of 1979-2010 from the twelve reanalysis products relative to the homogenized observations over China. The reanalysis products are (a) ERA-Interim, (b) NCEP-R1, (c) MERRA, (d) JRA-55, (e) NCEP-R2, (f) MERRA2, (g) ERA-20C, (h) ERA-20CM, (i) CERA-20C, (j) NOAA 20CRv2c, (k) NOAA 20CRv2 and (l) CFSR. The mainland of China is divided into seven regions (shown in Fig. 1c), specifically ① the Tibetan Plateau, ② Northwest China, ③ the Loess Plateau, ④ Middle China, ⑤ Northeast China, ⑥ the North China Plain and ⑦ South China.
Figure 2. The impact of inconsistencies between station and model elevations on the simulated multiyear-averaged differences in surface air temperatures ($T_a$, unit: °C) during the study period of 1979-2010 over China. The elevation difference ($\Delta \text{Height}$) between the stations and the models consists of the filtering error in the elevations used in the spectral models ($\Delta f$) and the difference in site-to-grid elevations ($\Delta s$) due to the complexity of orographic topography. $\Delta f$ is derived from the model elevations minus the ‘true’ elevations in the corresponding model grid cells from GTOPO30. The GTOPO30 orography is widely used in reanalyses, e.g., by ECMWF. The colour bar denotes the station elevations (unit: m). The relationship of the $T_a$ differences is regressed on $\Delta \text{Height}$ (shown at the bottom of each subfigure) or $\Delta f$ and $\Delta s$ (shown at
the top of each subfigure); the corresponding explained variances are shown.
Figure 3. Taylor diagrams for annual time series of the observed and reanalysed surface air temperature anomalies ($T_a$, unit: °C) from 1979 to 2010 in (a) China and (b-h) the seven subregions. The correlation coefficient, standard deviation and root mean squared error (RMSE) are calculated against the observed homogenized $T_a$ anomalies.
Figure 4. Composite map of partial correlation coefficients of the detrended surface air temperature ($T_a$, unit: °C) against surface incident solar radiation ($R_s$), precipitation frequency (PF) and surface downward longwave radiation ($L_d$) during the period of 1979-2010 from observations and the twelve reanalysis products. The marker ‘+’ denotes the negative partial correlations of $T_a$ with $R_s$ over the Tibetan Plateau in NCEP-R2, ERA-20C and ERA-20CM.
Figure 5. (a, b) The observed trends in surface air temperature ($T_a$, unit: °C/decade) and the simulated biases in the trends in $T_a$ (unit: °C/decade) during the period of 1979-2010 from (c) raw observations and (d-o) the twelve reanalysis products over China with respect to the homogenized observations. The squares denote the original homogeneous time series, and the dots denote the adjusted homogeneous time series. The probability distribution functions of all of the biases in the trends are shown as coloured histograms, and the black stairs are integrated from the trend biases with a significance level of 0.05 (based on two-tailed Student’s $t$-tests). The cyan and green stars in (k-n) represent estimates of the biases in the trends outside the ensemble
ranges whose locations are denoted by the black dots shown in \((k-n)\).
Figure 6. Composite map of the contributions (unit: °C/decade) of the biases in the trends in three relevant parameters, surface incident solar radiation \( (R_s, \text{ in red}) \), surface downward longwave radiation \( (L_d, \text{ in green}) \) and precipitation frequency (in blue) to the biases in the trends in surface air temperature \( (T_a) \) during the study period of 1979-2010, as estimated using the twelve reanalysis products over China.
Figure 7. Contribution (unit: °C/decade) of the biases in the trends in surface air temperatures ($T_a$) from three relevant parameters, surface incident solar radiation ($R_s$, in brown), surface downward longwave radiation ($L_{d}$, in light blue) and precipitation frequency (PF, in deep blue) during the study period of 1979-2010 from the twelve reanalysis products over China and its seven subregions.
Figure 8. Spatial associations of the simulated biases in the trend in surface air temperature ($T_a$) versus three relevant parameters among the twelve reanalysis products (solid lines indicate the NWP-like reanalyses, and dashed lines indicate the climate reanalyses). The probability density functions (unit: %) of these biases in the trends are estimated from approximately 700 1°×1° grid cells that cover China. The median values (coloured dots with error bars of spatial standard deviations) of the biases in the trends in $T_a$ (unit: °C/decade) in the twelve reanalyses are regressed onto those of (a) the surface incident solar radiation ($R_s$, unit: W·m$^{-2}$/decade), (b) precipitation frequency (unit: days/decade) and (c) the surface downward longwave radiation ($L_d$, unit: W·m$^{-2}$/decade) using the ordinary least squares method (OLS, denoted by the dashed grey lines) and the weighted total least squares method (WTLS, denoted by the solid black lines). The 5-95% confidence intervals of the regressed slopes obtained using WTLS are shown as shading. The regressed correlations and slopes are shown as grey and black text, respectively.