Sea Level Pressure Trends: Model-Based Assessment of Detection, Attribution, and Consistency with CMIP5 Historical Simulations

THOMAS R. KNUTSON[a] AND JEFFREY PLOSHAY[a]

[a] *Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, New Jersey*

(Manuscript received 27 December 2019, in final form 2 October 2020)

ABSTRACT: Observed sea level pressure (SLP) trends for 1901–10, 1951–10, and 1981–2010 are assessed using two observed data sources (HadSLP2_lowvar and 20CRv3) compared to a CMIP5 multimodel ensemble. The CMIP5 simulations include runs with (i) no external forcing (Control runs), (ii) natural external forcing only (Natural-Forcing), or (iii) natural plus anthropogenic forcings combined (All-Forcings). We assess whether the CMIP5 All-Forcing ensemble is consistent with observations and whether there is model-based evidence for detectable anthropogenic influence for the observed SLP trends. For the 1901–2010 and 1951–2010 trends, a robustly detectable anthropogenic signal in both observational data products is a zonal band of SLP increase extending over much of the Southern Hemisphere extratropics (30°–50°S). In contrast, the HadSLP2_lowvar and 20CRv3 observed data products disagree on the sign of the century-scale trends in SLP over much of the low-latitude region 25°N–25°S. These differences will limit confident detection/attribution/consistency conclusions for lower-latitude regions, at least until the observational data product discrepancies are better reconciled. The Northern Hemisphere extratropics remains a difficult region for identifying any detectable anthropogenic influence for annual- or seasonal-mean SLP trends. Overall, our results highlight the difficulty in detecting and attributing anthropogenic signals in SLP for relatively short time scales. The observed 1981–2010 regional trends typically have a different pattern and magnitude from the simulated externally forced trends. Consequently, our results suggest that internal variability is likely the dominant driver of most observed 1981–2010 regional trend features, including the pronounced increase in SLP over the central and eastern equatorial Pacific.

KEYWORDS: Pressure; Climate change; Climate models; Interdecadal variability; Trends

1. Introduction

This study explores how the atmospheric circulation may be changing regionally in response to anthropogenic forcing, based on an analysis of sea level pressure trends in observations and models. Previous studies have examined aspects of long-term trends and variability of various circulation indices and regional features. While not exhaustively reviewed here, previous investigations have explored Southern Hemisphere wind and SLP changes (e.g., Gillett et al. 2003, 2005, 2013; Gillett and Stott 2009; Swart and Fyfe 2012; Thompson et al. 2011); secular variability in the Walker circulation and associated sea surface temperature (SST) changes (e.g., Vecchi et al. 2006; Solomon and Newman 2012; Seager et al. 2019); changes in the Hadley circulation associated with a recent expansion of the tropics (e.g., Allen and Kovilakam 2017; Grise et al. 2018); changes in aspects of the Northern Hemisphere extratropical circulation, including waviness of the jet stream (Francis and Vavrus 2015; Coumou et al. 2018); and links between amplified Arctic warming or sea ice loss and Northern Hemisphere midlatitude circulation or extreme weather (Cohen et al. 2014, 2018; Blackport et al. 2019). Secular historical trend behavior of tropical cyclone metrics [reviewed in Knutson et al. (2019)] and Northern Hemisphere extratropical winter cyclone variability (Varino et al. 2019) have also been explored. In the present study, we focus specifically on regional trend behavior of the time-mean (annual and seasonal mean) circulation as inferred from SLP trends. We first document the observed SLP trends over roughly the past 110 years (1901–2010), 60 years (1951–2010), and 30 years (1981–2010). We then assess these observed SLP trends for evidence of detectable anthropogenic influence and their consistency with CMIP5 model historical simulated trends.

Confident detection of anthropogenic influence on atmospheric circulation has proven to be much more difficult than for global mean surface temperatures [e.g., Bindoff et al. (2013), in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5)]. This difficulty is apparent in a brief review of previous assessments of observed changes in the northern annular mode (NAM) and the related North Atlantic Oscillation (NAO). The NAM is an approximately zonally symmetric feature of variability of SLP in the Northern Hemisphere extratropics, while the NAO manifests as its regional counterpart over the Atlantic sector. The NAM and NAO are examples of Northern Hemisphere circulation (SLP) features that have received considerable attention concerning detection of possible anthropogenic influence. Noting the rising trend in the NAM index from the 1950s to the 1990s, the IPCC Fourth Assessment Report (AR4; Hegerl et al. 2007) concluded that this positive trend was likely related in part to anthropogenic forcing, although they noted that models simulated a smaller increasing trend than observed (i.e.,

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-19-0997.s1.

Corresponding author: Thomas R. Knutson, tom.knutson@noaa.gov

DOI: 10.1175/JCLI-D-19-0997.1

inconsistent with the observations). However, as the record extended further over time, a shorter-term decreasing tendency emerged that offset the longer-term increasing trend seen earlier. In the IPCC AR5, Hartmann et al. (2013), concluded that recent decreases in the NAM index had offset much of the longer-term increasing behavior, and that multidecadal variability was hampering the robust detection of long-term trends in the NAM and NAO. For the IPCC AR5 there was no anthropogenic attribution statement regarding the NAO/NAM. Multicentury reconstructions of the winter NAO index reviewed for the IPCC AR5 indicated that the observed twentieth-century changes in the index were not unprecedented in the context of these longer (~500 yr) reconstructions (Masson-Delmotte et al. 2013). A similar issue to the NAO/NAM case was noted in the AR5 for the Walker circulation (Hartmann et al. 2013) for which an observed weakening trend reported for the late nineteenth century to the 1990s had been largely offset by a more recent increase. These cases suggest the difficulties with inferring externally forced climate change responses from even relatively long-term circulation trends given the presence of potentially substantial multidecadal climate variability (e.g., Wunsch 1999).

Nonetheless, some anthropogenic influence on SLP has been formally detected at the global scale using pattern-based detection/attribution techniques, beginning with a seminal set of analyses by Gillett and collaborators. Gillett et al. (2003) had detected greenhouse gas and aerosol influence on December–February SLP changes, although they found that the models underestimated the magnitude of the observed changes. Gillett and Stott (2009) detected anthropogenic influence on SLP trends in other seasons and found that the most detectable changes were in low latitudes. Gillett et al. (2013) detected separate influences of greenhouse gases, aerosols, and ozone changes on SLP trends over 1951–2011. The IPCC AR5 (Bindoff et al. 2013) concluded that there was likely a detectable human influence on SLP patterns globally since 1951. However, the IPCC AR5 statement referred to global patterns, and so there were not specific attribution statements about SLP trends for individual regions (e.g., North America or Asia) that would be of interest for regional climate change applications.

The analysis of Gillett et al. (2013) is an example of formal pattern-based detection/attribution methodology in which observed changes are regressed onto a model-based pattern, or "fingerprint," of change. When the scaling coefficient for the regression between the model's fingerprint and observations is significantly greater than zero, climate change detection is claimed. When the coefficient is consistent with unity, attribution is claimed. In our study, we provide some additional information and assessment of observed SLP trends that is complementary to that of Gillett et al. (2013), notably for detection, attribution, and consistency at the regional scale. Using a regional univariate focus (but not a pattern-based focus), we assess trends at individual grid points to identify regions with consistency between a CMIP5 multimodel ensemble and observations, accounting for internal variability as estimated by the models. We examine three different periods (1901–2010, 1951–2010, and 1981–2010) to assess recent century-scale, 60-yr, and 30-yr trends. Similar to Gillett et al. (2013), we report where observed trends are unusual compared with model-simulated natural variability (they compared observations with simulated internal variability). We refer to trends that are highly unusual compared to natural variability as "detectable" as long as the trends are also in the direction expected from anthropogenic forcing. We also use our univariate approach to identify where an anthropogenic signal appears to be identifiable in the observations, based on where observed changes are both detectable and consistent with (or larger than) All-Forcing run trends. Our grid point–based consistency test simply assesses whether observed trends at a grid point are consistent or not with model All-Forcing runs, regardless of whether there is a detectable trend.

The pattern- or fingerprint-based approach used by Gillett et al. (2013) has certain advantages over the univariate approaches in that it makes use of the entire spatial pattern of change, rather than independently assessing behavior at different grid points. The combined use of space–time (i.e., multivariate) information provides a more comprehensive and powerful test for detection and attribution. However, the univariate approach can provide some useful complementary information, particularly for regional climate change assessment, as it is conceptually simpler than the pattern-based techniques. It does not rely on a model to supply a full pattern to be detected, but relies on the observations to define the patterns or global or regional changes to assess. When applied to individual regions, it can provide a useful initial assessment of detection, attribution, and consistency for observed regional trend features that may be of interest in that region. This type of information can be useful for regional climate change assessments. The univariate methodology has previously been applied for trends in surface temperature (Knutson et al. 2013), summertime wet bulb globe temperature (Knutson and Ploshay 2016), and precipitation (Knutson and Zeng 2018). It was used in the U.S. Fourth National Assessment Climate Science Special Report (Vose et al. 2017) to assess and detect anthropogenic influence on surface temperature trends over much of North America region at the grid point (5° × 5°) scale. Bhend and Whetton (2013) used a similar univariate method to assess individual models for consistency with observed SLP trends for 1956–2004, although they were not assessing the trend for regional detection of anthropogenic influence. Using the univariate approach in our study, we can re-examine some regional circulation trend features that have been identified in previous publications. Such an analysis can thus provide a useful updated perspective on regional observed trends, including their consistency with model simulations, as well as an assessment of whether and where SLP trends provide evidence for detectable human influence on atmospheric circulation.

In the remainder of this report, we describe the univariate detection, attribution, and consistency methodology in more detail in section 2, along with some related methodology. The main assessment results are then presented in sections 3–6, and the summary and conclusions are presented in section 7.

2. Methodology

The univariate methodology used for the trend assessment can be divided into three basic tasks: 1) calculation of the

observed and modeled trends at each grid location with adequate observed data coverage for the trend analysis, using the observed data mask to artificially limit coverage in space and time for the model simulations; 2) development of trend distributions for the Natural-Forcing and All-Forcing simulations using a set of CMIP5 models from different modeling centers; and 3) categorization of observed trends in terms of consistency, detection, and attribution. These steps are described in more detail below.

a. Calculation of observed and model trends

The HadSLP2_lowvar dataset on a 5° × 5° grid (Allan and Ansell 2006; see https://www.metoffice.gov.uk/hadobs/hadslp2/data/download.html) was used for our initial trend analysis where we considered data from 1901 to 2010. This dataset comprises the original HadSLP2 data through 2004, combined with an extension through 2012 based on NCEP–NCAR reanalysis data, adjusted to have variance levels consistent with the HadSLP2 data (termed "HadSLP2_lowvar" here; see https://www.metoffice.gov.uk/hadobs/hadslp2/data/download.html). As an alternative observationally based dataset, we tested version 3 of the NOAA-CIRES-DOE Twentieth Century Reanalysis (20CRv3; Slivinski et al. 2019). The 20CRv3 uses an atmospheric model with specified SSTs and sea ice, which is run over historical, time-evolving SSTs and sea ice. Into this time-evolving system, sea level pressure is the only piece of observational atmospheric data that is assimilated. The model then simulates the time-evolving behavior of the remaining fields as the assimilation proceeds over the historical period. While the 20CRv3 has no missing data, the model is effectively infilling missing data during the integration. The HadSLP2_lowvar dataset includes missing data flags for months and locations without data. For our study, since we were interested in an intercomparison of these two data sources with regard to trends, we masked out the 20CRv3 data for those time periods and locations where the HadSLP2_lowvar dataset reports missing data. While this is not an ideal screening of the 20CRv3 for inadequate data coverage, as there are differences in SLP input data for the two observational products, our approach facilitates comparisons between the two datasets over a common period (1901–2010) with regard to longer-term trend behavior.

In order for an annual mean trend to be computed at a given grid point using either (masked) dataset, we screened the data using a temporal coverage threshold requirement. Adequate coverage for a time period (e.g., 1901–2010, 1951–2010, or 1981–2010) was defined as having at least 20% data availability for both the first 10% and last 10% of the full period, as well as 70% overall data availability.

For our analysis, all model data as well as the 20CRv3 data were initially regridded onto the HadSLP2_lowvar dataset's grid using Ferret (https://ferret.pmel.noaa.gov/Ferret/), before masking the 20CRv3 or modeled data with the HadSLP2_lowvar observed data mask. For the CMIP5 model control runs, segments at random starting times were obtained by assigning the first year of the model control run segment to the initial year of a trend period (1901, 1951, or 1981), and all control run segments used for trend analysis were also masked with the HadSLP2_lowvar observed mask of missing data periods/grid points.

b. Construction of CMIP5 Natural-Forcing and All-Forcing trend distributions

A total of 10 CMIP5 models were used in our analysis. The 10 models were chosen [as in Knutson and Zeng (2018)] because they had available both Natural-Forcing and All-Forcing runs that extended to at least 2010 in the CMIP5 database. The All-Forcing runs include both anthropogenic (e.g., greenhouse gases, aerosols, land-use change) and natural (e.g., solar variability, volcanic aerosol loading) forcing agents, while the Natural-Forcing runs contain only natural forcing agents (e.g., solar variability and volcanic aerosol loading). Both All-Forcing and Natural-Forcing runs also contain internal climate variability, and the internal climate variability characteristics of each model can be estimated from long (multicentury to millennial-scale) preindustrial control simulations that have no changes in anthropogenic or natural forcing agents over time. This assumes that the internal variability does not change systematically with climate change over the period of study, so that it can be represented throughout by the preindustrial control run variability.

For each individual CMIP5 model, ensemble mean trend maps were computed by averaging the trends across available ensemble members (Natural-Forcing or All-Forcing). Then these individual model ensemble mean trend fields were averaged to obtain the multimodel grand ensemble mean trend field maps. Examples of these modeled trend field maps for historical runs and for twenty-first-century projections are shown for various periods in Fig. 1 (without data masking). These represent the average expected signal in SLP trends from the 10 CMIP5 models for the various past and future time periods, and assuming no missing data. Specifically, Fig. 1 shows the CMIP5 ensemble mean SLP trends from the All-Forcing historical simulations (1901–2010, 1951–2010, and 1981–2010) and the twenty-first-century projected trends (2006–2100) under the RCP8.5 forcing scenario (Fig. 1d). The historical simulation results were extended from 2006–10 using the RCP8.5 forcing scenario, although results are similar for other RCP scenarios over this short time period.

The main feature in the CMIP5 modeled trends for 1901–2010 (Fig. 1a) is a couplet of increasing SLP trends along 35°–50°S, accompanied by decreasing trends in a circumpolar region extending from about 60°S to the South Pole. Weaker negative trends in SLP are modeled over most of the northern extratropics, and in parts of the Southern Hemisphere tropics and subtropics, including the southeastern Pacific. As shown by Gillett et al. (2013) the Southern Hemisphere extratropical SLP couplet is a characteristic response of climate models to a combination of greenhouse gas increases and stratospheric and tropospheric ozone changes, maximizing during the austral summer season (see their Fig. 1, DJF column).

For the 1951–2010 period (Fig. 1b) the CMIP5 models hindcast a relatively smoothly structured pattern for SLP trends, similar to that for 1901–2010 but with higher absolute trend magnitudes, consisting of a zonally oriented couplet of increasing and decreasing trends in the extratropical Southern Hemisphere, and generally weaker decreasing SLP trends in the northern extratropics.

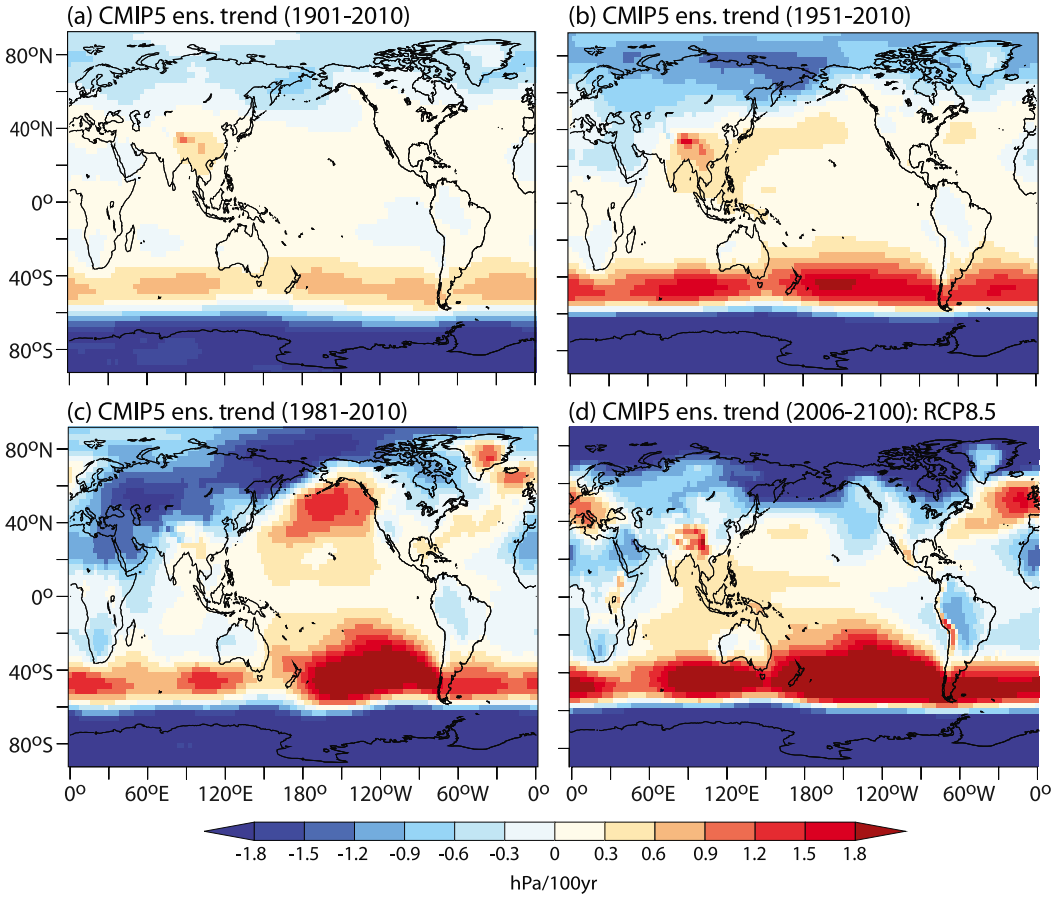

FIG. 1. Ensemble mean SLP trends [hPa (100 yr^{-1})] for various historical and future time periods, including All-Forcing historical simulations for (a) 1901–2010, (b) 1951–2010, and (c) 1981–2010, and (d) projections for 2006–2100 under the RCP8.5 forcing scenario.

For 1981–2010 (Fig. 1c), the modeled trend map again shows the Southern Hemisphere extratropical couplet, with increasing SLP trends in the midlatitude Southern Hemisphere, decreasing SLP trends in the high-latitude Southern Hemisphere, and weaker general decreases in SLP over Africa, Eurasia, and the Arctic. A unique feature in the modeled ensemble trend field for 1981–2010, compared to 1901–2010 or 1951–2010 trends, is the pair of lobes of positive SLP trends in the north and south extratropical Pacific. The northern extratropical positive trend feature in the Aleutian low region does not appear in the CMIP5 RCP8.5 projection trends (Fig. 1d), nor in the 1901–2010 trend field (Fig. 1a). Examination of individual model time series for this region (not shown) suggests that this feature is mainly due to residual internal variability owing to the modest sample size of forced simulations available, the short period of the trend, and relatively high internal variability in the region. The South Pacific lobe feature in Fig. 1c has some expression in the 2006–2100 projection and so may be an emerging zonal asymmetry in the SLP response to anthropogenic forcing.

Our assessment methodology requires not just the mean trends from the models but an assessment of the *expected distribution* of trends that can be expected on various time scales due to internal climate variability. To address this issue, the 5th to 95th percentile range of trends was estimated for each model by sampling (with replacement) appropriate-length segments from that model's Control run (i.e., multicentury simulation with no external forcing changes). We note that it would be advantageous to have either a very large set of Natural Forcing and All Forcing ensemble members or very long (multimillennial) control runs from all modeling centers in order to more robustly estimate the 5th and 95th percentiles of the 30-, 60-, and 110-yr trends in the Control runs. Our estimates of these percentiles are still relatively crude given the limited simulation lengths available.

We then combined trend samples from the 10 models to construct multimodel ensemble distributions of All-Forcing and Natural-Forcing trends. This construction can be done in various ways, and we have created ensemble distributions using two different methods for comparison. Our methods are illustrated in Fig. 2 for the example case of SLP anomalies averaged over the Southern Hemisphere extratropical band 30°–50°S. This region encompasses the region of hindcast and projected increasing SLP trends for the Southern Hemisphere

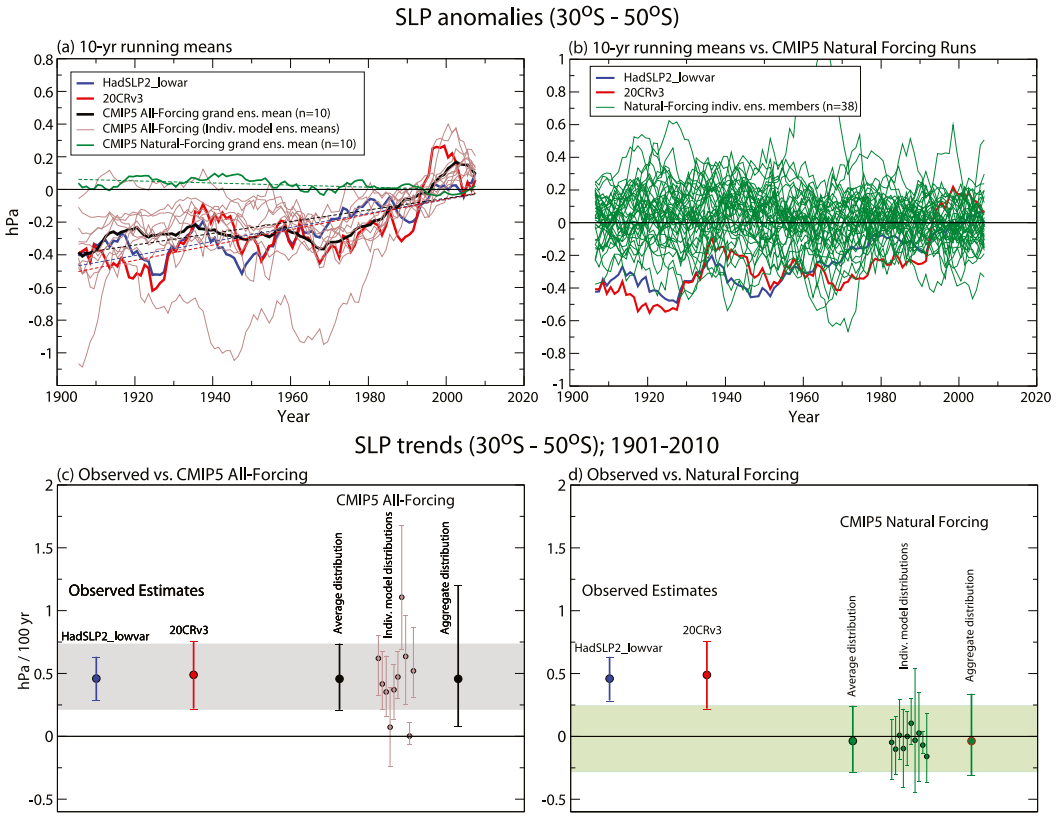

FIG. 2. Illustration of methodology for constructing multimodel ensemble All-Forcing and Natural-Forcing trend distributions for SLP anomalies for 30°–50°S. (a) 10-yr running mean anomalies for observations (HadSLP2_lowvar; 20CRv3), All-Forcing ensemble mean, and Natural-Forcing ensemble mean (see legend), relative to 1981–2010 base period. Thin gray lines depict 10 individual model ensemble means for All-Forcing runs from the 10 models. (b) 10-yr running means from 38 Natural-Forcing individual ensemble members across all 10 models (green lines) along with observed SLP anomalies from HadSLP2_lowvar and 20CRv3, as in (a). (c) Means (dots) and 5th–95th percentile ranges (bar/whiskers) of trends [1901–2010; hPa (100 yr^{-1})] from observations (HadSLP2_lowvar, 20CRv3) and CMIP5 All-Forcing simulations. Thick black dots/bars show average and aggregate 10-model distributions (see text for details), with the distributions (bars) based on control run trends sampled from the 10 models. Individual model trend distributions (mean and 5th/95th percentiles) are shown by small dots and gray bars. (d) As in (c), but for the CMIP5 Natural-Forcing simulations, depicted by green dots/bars. Gray and green shaded bands in (c) and (d) denote the 5th–95th percentile ranges of the All-Forcing and Natural-Forcing average distributions (see text).

extratropics shown in Fig. 1. The observational estimates (HadSLP2_lowvar and 20CRv3) for this region show a rising trend (Fig. 2a) that is generally distinct from the Natural-Forcing simulations in terms of century-scale behavior (Fig. 2b) but relatively consistent with the typical All-Forcing model simulations (Fig. 2a). These comparisons are further quantified in Figs. 2c,d as explained below.

The first ensemble method we used to create an ensemble trend distribution is the *average* model distribution method, where the mean of the trend distribution is defined by the grand ensemble mean of the 10 models' trends (either All-Forcing or Natural-Forcing), while the 5th and 95th percentiles were defined by this mean combined with the average of the 5th/95th percentiles of sampled trends across the 10 model control runs. These average trend distributions (All-Forcing or Natural-Forcing) then represents the behavior of a single hypothetical model having the average trend characteristics of the multimodel ensemble for those forcing assumptions. These are shown in Figs. 2c,d by the gray shading (All-Forcing hypothesis) and green shading (Natural-Forcing hypothesis), respectively, where they can be compared to the observed trends from HadSLP2_lowvar and 20CRv3.

Alternatively, we constructed multimodel *aggregate* Natural-Forcing and All-Forcing trend distributions, where the sampled trends were computed by combining each model's All-Forcing or Natural-Forcing ensemble-mean trend with large samples of that model's internal variability trends as obtained from its control run. Results from all 10 models were then combined into a single aggregate distribution, and the mean, 5th, and 95th percentiles of this aggregated distribution were computed. The aggregate trend distributions will generally have a larger spread than the average distribution owing to

the spread contributed by the differences in mean trends across the 10 models (Figs. 2c,d).

Note that both the average and aggregate distributions of trends included randomly selected (resampled with replacement) control run trends with an equal number of samples being drawn from each individual model. Thus each of the 10 models had equal influence on the distribution regardless of its control run length. The average model distribution method was our primary calculation, while the aggregate distribution method was used as a sensitivity test; some sample comparison assessment results and discussion are contained in the online supplemental material. These methods are discussed further in Knutson et al. (2013). The comparison of trends in our example case for 30°–50°S SLP (Figs. 2c,d) indicates that for the 1901–2010 period, the observed trend estimates (blue- and red-filled circles for HadSLP2_lowvar and 20CRv3, respectively, in Figs. 2c,d) are inconsistent with the average and aggregate Natural-Forcing trend distributions, but consistent with the average and aggregate All-Forcing simulations. That is, the central estimates for the observed trends are outside the 5th–95th percentiles of the Natural-Forcing distributions but inside the 5th–95th percentiles for the All-Forcing distributions.

c. *Assessment of modeled internal variability*

The internal variability as obtained from model control runs is an important metric for our assessment as it provides an estimate of the expected internal variability in the real world. However, model control run variability is not directly comparable to observations. In contrast to model Control runs, observed variability contains both internal (natural intrinsic) variability, along with forced natural variability caused by natural external forcings (solar variability and volcanic forcing), as well as forced anthropogenic variability caused by anthropogenic forcings. To reduce the effects of this difference, we construct a modified version of observations for comparison to the models' control run variability (Fig. 3). We first remove an estimate of the external forced response from observations (either HadSLP2_lowvar or 20CRv3) to form an observed internal variability estimate. To do this, we subtract the CMIP5 multimodel ensemble All-Forcing response from the observations and compare the standard deviations of the residual observed time series to the standard deviations from the model Control runs at each grid point. We restricted the period analyzed to 1951–2012 in order to have relatively broad spatial coverage.

These grid point–based internal variability comparisons for the case of HadSLP2_lowvar (Figs. 3a,b,e) suggest that the CMIP5 models on average tend to underestimate SLP variability (blue colors in Fig. 3b) in tropical latitudes and overestimate it (red colors) in extratropical latitudes. Thus a caveat of our study is that the CMIP5 models may bias climate change detection results such that we overestimate the detection of unusual (compared to internal variability) SLP trends in tropical latitudes but underestimate it in the extratropics, when comparing to HadSLP2_lowvar observations.

Comparing the modeled control run variability to the 20CRv3 estimate of observed internal variability (Figs. 3c–e) provides a different result. In this case blue colors predominate in Fig. 3d, indicating that the modeled internal variability tends to be smaller than the observed estimate, especially in the tropics and subtropics. A caveat of the model–observation variability comparison for 20CRv3 is that the observed data product has much larger SLP trends than observed in the tropics and subtropics (which will be shown and discussed later), which can inflate the residual variability estimate that we use for comparison to the control runs, and is likely responsible for at least some of the discrepancy shown in Fig. 3d.

Note that we have not focused on the low-pass (>10 yr) variability for the variability comparisons in Fig. 3 due to the relatively short period of observations available over much of the globe. Low-pass variability, or ideally a very large sample of multidecadal to century-scale trends from observations, would be preferable to compare to the modeled distribution of these same metrics, but this is not possible due to the short period of observations, so the interannual variability we test is a crude metric of variability for the purposes of evaluating modeled low-frequency variability for our study. For long-term (e.g., century scale) trend assessment, it would also be advantageous to have much larger samples of century-scale trends from models than are currently available, as discussed earlier.

d. *Categorization of trends for consistency, detection, and attribution*

Comparing the observed trend to the multimodel All-Forcing and Natural-Forcing distributions of trends described above, we classified the observed trends at each grid point into assessment categories. These are used to produce our categorization maps shown for example in Fig. 4b, for which the different categories (colors) are explained further below. The All-Forcing and Natural-Forcing trend distributions have identical internal climate variability contributions, as these were obtained from the long-running preindustrial Control runs according to our methodology. We view the All-Forcing runs, in short, as representing the best attempt by each modeling center to simulate the statistical distribution of historical forced trends, while the difference between the All-Forcing and Natural-Forcing distributions represents the estimated anthropogenic influence.

We followed the assessment categorization methodology described by Knutson and Zeng (2018) for precipitation trends (see visual summary in their Fig. 2), which is briefly reviewed here. *Consistency* is defined as where the observed trend lies within the 5th–95th percentile range of the multimodel ensemble distribution of trends from the All-Forcing runs, and is denoted by white stippling on various colored backgrounds in Fig. 4b. *Detection* of a climate change is defined as having the observed trend lie outside the 5th–95th percentile range of the Natural-Forcing distribution, and outside the 5th–95th percentile range of the Control distribution, with the additional constraint that the observed trend should have the same sign as the inferred anthropogenic influence (which is inferred from comparison of the All-Forcing and Natural-Forcing run ensemble mean trends). In our categorization maps (e.g., Fig. 4b) nondetection is shown by gray, orange, or green shading. Gray shading denotes where the trend is within the 5th–95th percentile distribution of either the Natural-Forcing or Control

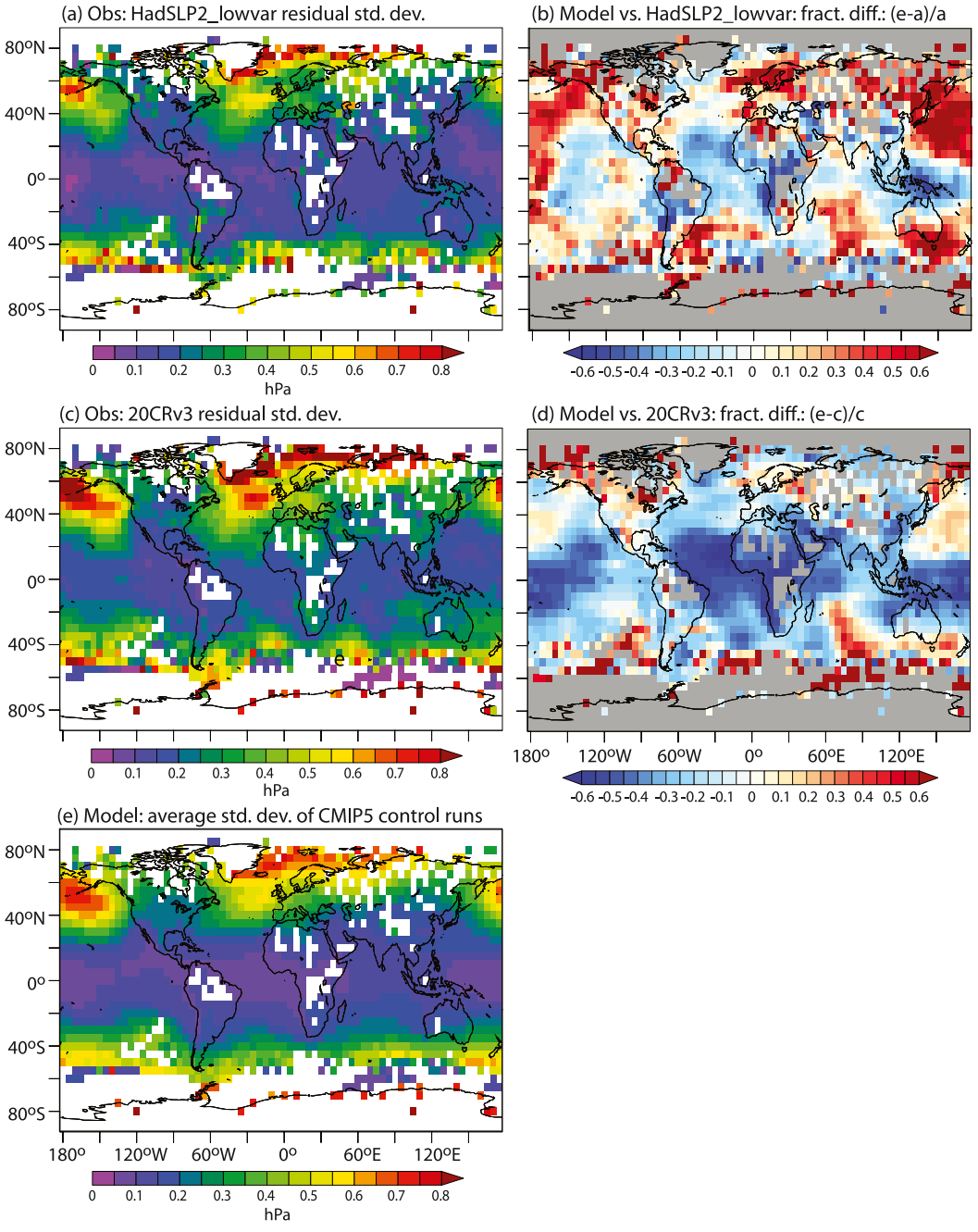

FIG. 3. Standard deviation of annual mean SLP anomalies (hPa) for (a) HadSLP2_lowvar or (c) 20CRv3 observed data products (1951–2012), where an estimate of the externally forced response (i.e., the CMIP5 multimodel ensemble All-Forcing response) was subtracted from each observed product before computing the standard deviation; and (e) the multimodel average of the standard deviations of interannual values of SLP from the CMIP5 model control runs. (b),(d) Fractional difference maps defined by (observations − model)/observations. For (b), this is [(e) − (a)]/(a); for (d), this is [(e) − (c)]/(c).

run distribution. The case of having an observed trend outside the 5th–95th percentile range of the Natural-Forcing and Control run distributions (i.e., large in magnitude), but having the opposite sign as the All-Forcing grand ensemble mean trend, was not classified as *detectable or nondetectable*, but rather as a *large unexplained* trend (orange or green shading in Fig. 4b). At least partial *attributable anthropogenic influence* was inferred if the observed trend was both inconsistent with (outside 5th–95th percentile range of) the Natural-Forcing and Control run distributions, and consistent with (within the

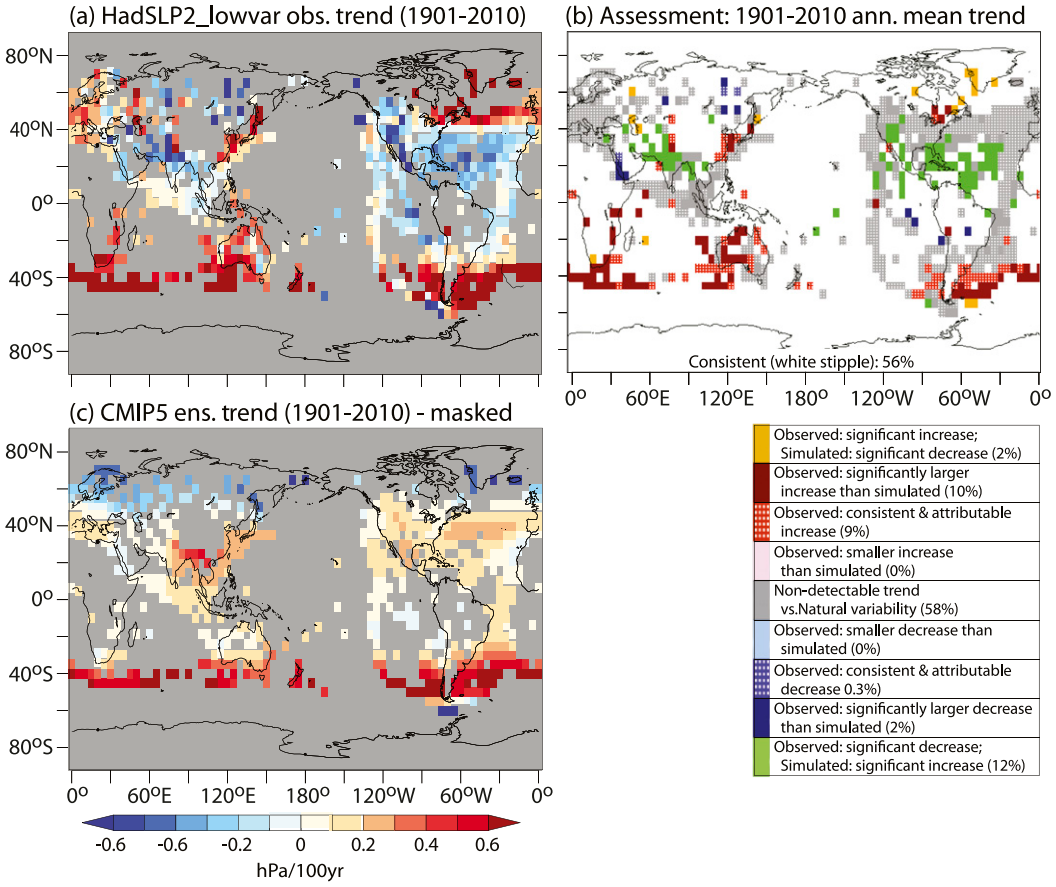

FIG. 4. Assessment of observed (HadSLP2_lowvar) trends in annual mean sea level pressure (SLP) over 1901–2010 based on comparison with an ensemble of 10 CMIP5 models. (a) Observed (HadSLP2_lowvar) and (c) CMIP5 multimodel ensemble trends [hPa (100 yr^{-1})]. (b) Model-based assessment of the observed trend at each grid point having sufficient data coverage for trend analysis (see text). Nine assessment categories are defined (see color scale, legend, and text for details). The percent of analyzed area classified in each category is listed in parentheses in the legend box. Grid locations where the models' All-Forcing simulation time series are consistent with the observed trend are identified with white stippling. Regions with inadequate data coverage over 1901–2010 are denoted by gray shading in (a) and (c) and by white shading in (b). A small fraction of grid points that did not fit within any of our defined categories are also shaded white in (b).

5th–95th percentile range of) the All-Forcing distribution (e.g., blue-stippled or red-stippled shaded regions in Fig. 4b). An additional set of conditions that we also inferred as demonstrating at least some detectable and attributable anthropogenic influence occurred in the case where the observed trend was both inconsistent with the Natural Forcing and Control run distributions and also extended beyond the 5th or 95th percentile range of the All-Forcing distribution (while also being in the same direction as the anthropogenic influence inferred from models by comparing the All-Forcing and Natural Forcing mean trends). In other words, some detectable anthropogenic influence is also inferred if the observed trend is highly unusual compared to natural variability estimates and is even much bigger than the modeled All-Forcing response, while at least the modeled anthropogenic response and the observed trend are in the same direction (e.g., solid red or solid blue shading in Fig. 4b).

We note that inconsistencies between observed trends and the modeled All-Forcing run trends (nonstippled regions in Fig. 4b) can be due to four types of errors. These include (i) errors in model-simulated responses to the specified climate forcing changes, (ii) errors or omissions in the specified forcing changes, (iii) possible error (underestimation) of internal climate variability in the models, and (iv) errors in observations leading to errors in the observed trend field. Thus our results cannot strictly be used to falsify a model because of ambiguity about the relative contributions of the above four factors to any discrepancies noted between All-Forcing runs and observations. Disentangling these various potential causes of discrepancies remains a substantial research challenge that is beyond the scope of the present study.

3. Trend assessment: HadSLP2_lowvar

In this section we present summary maps showing the trend assessment results at the grid point scale for three different time periods: 1901–2010, 1951–2010, and 1981–2010.

a. 1901–2010 trends assessment: HadSLP2_lowvar

The HadSLP2_lowvar observed trends for 1901–2010 are shown in Fig. 4a. As indicated in the figure, the observed coverage for trend assessment is relatively sparse but with enough coverage to show at least a very broad-scale view of the SLP trend field over the 110-yr period. The trend field shows a zonally oriented band of increases in SLP spanning the Southern Hemisphere near 40°S. Similar increases are also seen in the CMIP5 All-Forcing ensemble mean response (Fig. 4c) although the model ensemble-mean trends are relatively weaker than the observed trends. The main features in the unmasked modeled trends in SLP for 1901–2010 (Fig. 1a) were discussed previously. Due to the poor sampling in observations, many of the modeled trend features in the (unmasked) historical runs (Fig. 1a) cannot be compared confidently to observations.

For the trend assessments (Fig. 4b), we plot and assess trends only for those grid points with sufficient historical coverage for trend analysis (as defined in the methodology). Based on these assessed regions, the fraction of analyzed area where the All-Forcing ensemble trend distribution is consistent with the observed trend is 56%. The summary trend assessment further indicates that the observed Southern Hemisphere extratropical increases are detectable trends (unusual compared to the natural variability), and that for many grid points in the region the positive observed trends are larger than even the 95th percentile of the All-Forcing distribution, meaning the trends are relatively weak in models compared to the much stronger trends in observations. This strong detectable and partly attributable observed Southern Hemisphere increasing SLP trend signal is consistent with previous findings (e.g., Gillett et al. 2013). Figure 4b suggests that detectable and at least partly attributable anthropogenic increases occur over about 19% of the analyzed area. There is only a minor area (less than 3% of analyzed area) with detectable decreases (blue boxes) globally. There are also quite a few regions where modeled and observed SLP trends have opposite signs (cf. Figs. 4a and 4c). Among the most notable of these is the 12% of analyzed area with green boxes in Fig. 4b, where observed decreases are relatively strong and are below the 5th percentile of the Natural-Forcing distribution; however, they are not classified as detectable because the All-Forcing simulation shows an increase in contrast to the observed decreases. These "unexplained decreasing trend" regions include parts of the northern tropical and subtropical Atlantic, eastern Pacific, North America, and South Asia.

The summary trend categorization (Fig. 4b) also indicates that the observed trends were nondetectable compared to Natural-Forcing and Control runs over 58% of the analyzed domain (gray regions), reinforcing how challenging detection/attribution of SLP grid point–based trends is even at the century scale, compared with surface temperature, for example (Knutson et al. 2013, see their Fig. 10). A few additional minor features in the analysis include a suggestion of a possible region of anthropogenic decrease in SLP represented by scattered grid points over extratropical northeastern Asia. As shown by the unmasked version of modeled trends (Fig. 1a), there is a modeled expectation for only weak decreasing trends in SLP over 1901–2010 across most of the Northern Hemispheric middle to high latitudes, which is qualitatively consistent with the relatively rare occurrence of detectable trends there in our assessment.

As discussed in the methodology section (e.g., Fig. 2) for the CMIP5 Natural-Forcing and All-Forcing trend distributions, our principal methodology is to compute and compare observed trends to an *average* distribution of trends from the model All-Forcing or Natural-Forcing runs, which are based on the control-run-estimated average spread in trend distribution due to internal variability. Using the alternative method of the *aggregate* combined distribution of trends from all models (rather than their multimodel average) gives wider trend distributions from the model ensemble. In the online supplemental material (see Fig. S3) we show a sensitivity test—assessment maps for 1901–2010 trends directly comparing results using the aggregate and average methods. As expected, using the aggregate distribution method we find a greater fraction of area with consistency between All-Forcing runs and observations than using the average distribution method, but we also find a slightly smaller fraction of area with trends that are distinguishable from the Natural-Forcing trends. However, the overall results and main features of the trend assessment are fairly similar using the average and aggregate distribution methods.

b. 1951–2010 trends assessment: HadSLP2_lowvar

For 1951–2010 trends (Fig. 5), the observed area coverage for our trend assessment is increased relative to 1901–2010 (Fig. 5a vs Fig. 4a). In HadSLP2_lowvar observations (Fig. 5a), a major feature is a large zonally oriented region of increasing trends in the Southern Hemisphere extending from the tropics to about 40°S, similar to the simulated feature in the model All-Forcing runs over this region (Fig. 5c). This is the most pronounced inferred detectable anthropogenic influence for 1951–2010 SLP trends (Fig. 5b) and is relatively consistent with the results for 1901–2010. These general features in Figs. 5a and 5c are also evident in Gillett et al.'s (2013) seasonal trend analysis (their Fig. S2). One distinction is that our annual mean trend assessment shows little evidence for detectable trends over the tropical Pacific sector, whereas Gillett et al. find that observed 1951–2011 seasonal increasing trends are unusual compared to control run variability over much of the tropical Pacific during most seasons, but with unusually strong *decreasing* trends there during the SON season. Our results are qualitatively consistent with Swart and Fyfe (2012), who reported that historical trends in surface westerlies since 1979 and since 1951 in the Southern Hemisphere surface consist primarily of a strengthening of the westerlies.

The fraction of analyzed area where the All-Forcing ensemble trend distribution is consistent with the observed trend is 80% (Fig. 5b), which is a substantially greater fraction than for 1901–2010. However, a greater fraction of analyzed area also has only weak nondetectable trends, consistent with Natural-Forcing runs; for 1951–2010, this is 80% of analyzed area compared to 58% for 1901–2010.

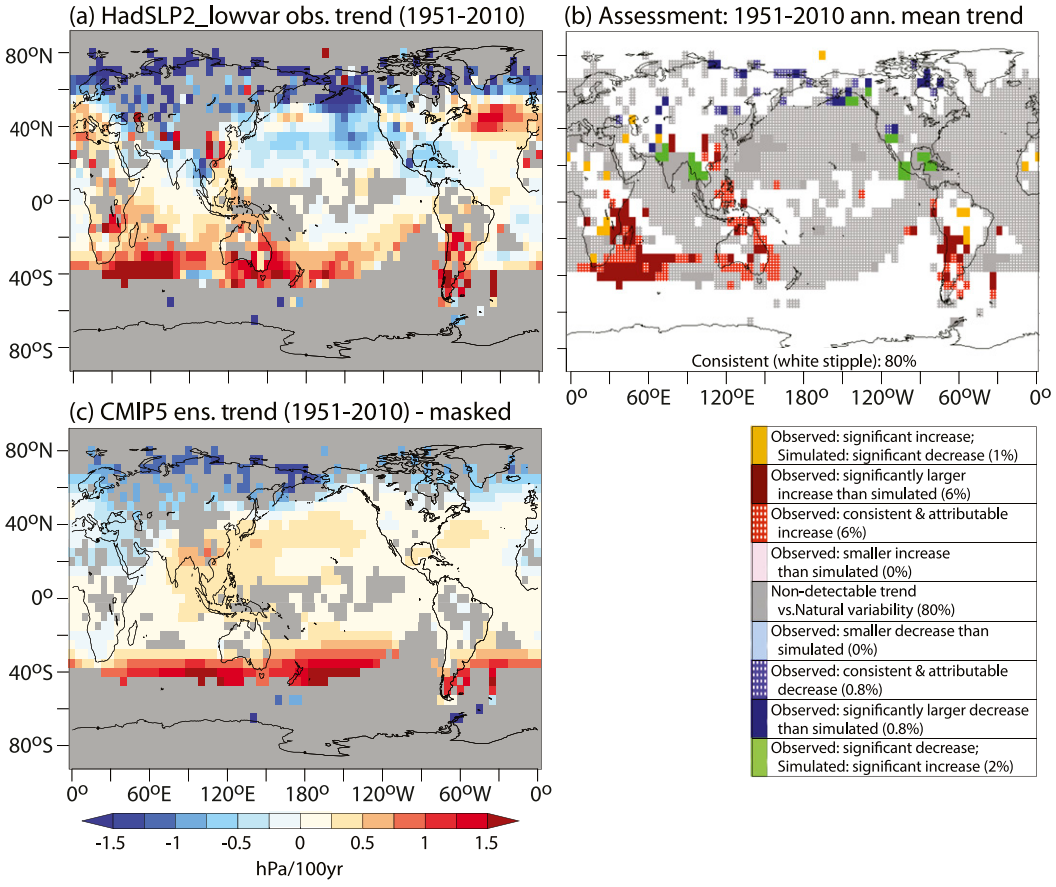

FIG. 5. As in Fig. 4, but for 1951–2010 trends.

The scattered blue boxes across the middle and higher latitudes of the Northern Hemisphere are suggestive of an emerging detectable decrease in SLP over broad regions of the northern extratropics. As suggested by the unmasked model data (Figs. 1a,b) this is a feature that model hindcasts project would be stronger for 1951–2010 trends than for 1901–2010 trends, as hinted at by the limited observations. However, we find that only about 2% of analyzed area has detectable decreases in SLP over 1951–2010, indicating the relative weakness of this signal, despite its spatially coherent structure. Trend in the northern extratropics area-average SLP will be discussed in section 6.

The categorization map (Fig. 5b) also indicates significant (outside the 5th to 95th percentile range of the Natural-Forcing distribution) but unexplained decreases in SLP (green shading) in a few low-latitude regions (2% of area) mainly in southern North America, the Gulf of Mexico, and parts of South and Southeast Asia.

c. 1981–2010 trends assessment: HadSLP2_lowvar

For 1981–2010 trends (Fig. 6), the observed coverage for trend assessment is much improved relative to the longer-period analyses (Figs. 4 and 5). The modeled All-Forcing trend (Fig. 6c) includes only relatively small increasing SLP trends across the tropical Pacific, while the observed trend (Fig. 6a) comprises a more pronounced increase over the eastern tropical Pacific and a decrease in the western tropical Pacific and over the Maritime Continent (opposite in sign to the modeled trend there). These observed trend features in the tropical Pacific, indicative of an increase in the Walker circulation over this period, are highly unusual compared to the model's Natural-Forcing distribution. However, since the model simulates the wrong direction of trend in the western Pacific/Maritime Continent regions, the only detectable and partly attributable signal within the Pacific sector is the broad region of increasing SLP trends in the eastern tropical Pacific and subtropical southeast Pacific. Within that region the observed trend is larger than the 95th percentile of the All-Forcing trend distribution (e.g., the dark-red regions of Fig. 6b). The assessment map (Fig. 6b) further identifies some regions of detectable and attributable decreasing trends (blue regions) over parts of central Asia, Africa, the Middle East, and the eastern Atlantic. In some of this region, particularly over the Atlantic Ocean, the observed decreasing trend is substantially stronger than (below the 5th percentile of) the All-Forcing trend distribution.

The increasing SLP trend feature in the tropical Pacific (Fig. 6b) is consistent with the recent increase in the Walker circulation reported in IPCC AR5 (Hartmann et al. 2013) and will be discussed further below in section 6.

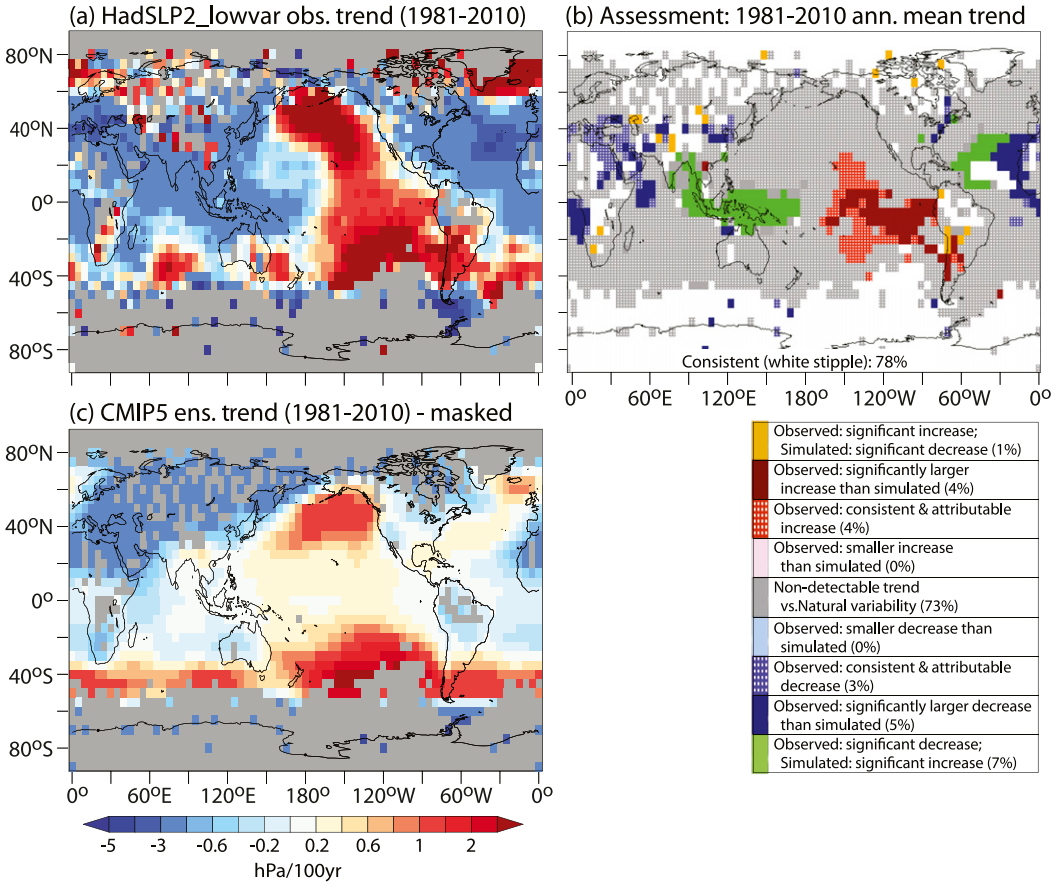

FIG. 6. As in Fig. 4, but for 1981–2010 trends.

For the extratropics, there are almost no detectable trends identified poleward of about 40°N/S (Fig. 6b). Overall, 73% of the analyzed area globally had trends over 1981–2010 that were classified as not unusual compared to the Natural-Forcing and Control run distributions (i.e., nondetectable). For example, the notable increasing trend feature in the observations for the Aleutian low region (Fig. 6a) is not classified as detectable according to our analysis of annual means. This seems qualitatively consistent with Sun et al. (2019), who report on statistically significant Aleutian low weakening during springtime (1980–2013). They attributed the observed change primarily to internal variability, noting that external forcing seems to be playing just a minor role in the changes. In our analysis (Fig. 6b), despite the modest fraction of area with detectable changes, most of the analyzed area (78%) at least had changes that were consistent with the model All-Forcing simulation ensemble.

4. Alternative observational data product: 20CRv3

In this section, we explore the sensitivity of our trend analysis to the use of an alternative observational data product—the 20CRv3 (Slivinski et al. 2019). For a first comparison of the long-term trend behavior between the HadSLP2_lowvar and 20CRv3, Fig. 7 shows the difference (HadSLP2_lowvar minus 20CRv3) between annual mean SLP anomalies of the two datasets for the period 1901–30. For both datasets, anomalies were first created for the entire period of record relative to a 1981–2010 reference period. The comparison thus shows that 20CRv3 has lower or more negative early twentieth-century SLP anomalies in comparison with HadSLP2_lowvar over most tropical latitudes (particularly over the Pacific and Atlantic sectors). On the other hand, in the extratropical Southern Hemisphere, from about 40°S to 60°S, HadSLP2_lowvar had negative SLP anomalies in comparison to 20CRv3. The map shows that the differences between HadSLP2_lowvar and 20CRv3 are coherent on a relatively large scale, with the aforementioned differences being the most prominent features.

To explore these differences in Fig. 7 further, time series of SLP averaged over 25°N–25°S are compared for the HadSLP2_lowvar and 20CRv3 datasets in Fig. 8. The two datasets disagree remarkably on the nature of the tropical long-term trend, with HadSLP2_lowvar having a generally negative trend, especially over the first half of the record, and 20CRv3 having a pronounced positive trend, especially in the second half of the record. We fit a cubic least squares curve to both datasets to remove most of the slowly evolving component and an overall difference in mean values, so that the higher-frequency variability could be more easily compared. The remaining residual

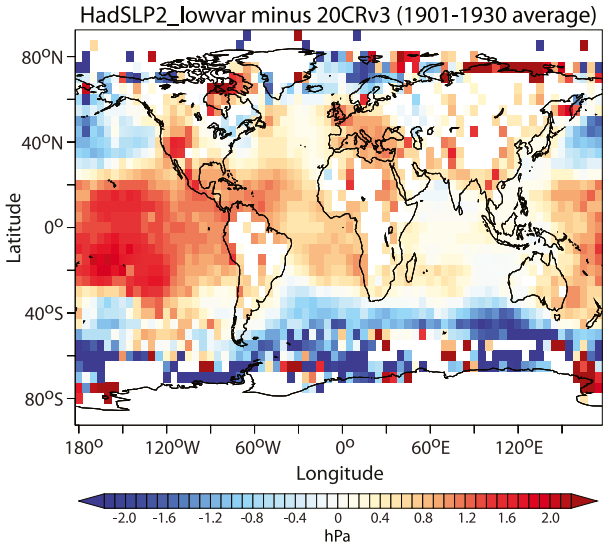

FIG. 7. Difference between two observational products (HadSLP2_lowvar minus 20CRv3) for annual mean SLP anomalies (hPa) averaged over 1901–30. The anomalies in both cases are relative to a 1981–2010 reference period.

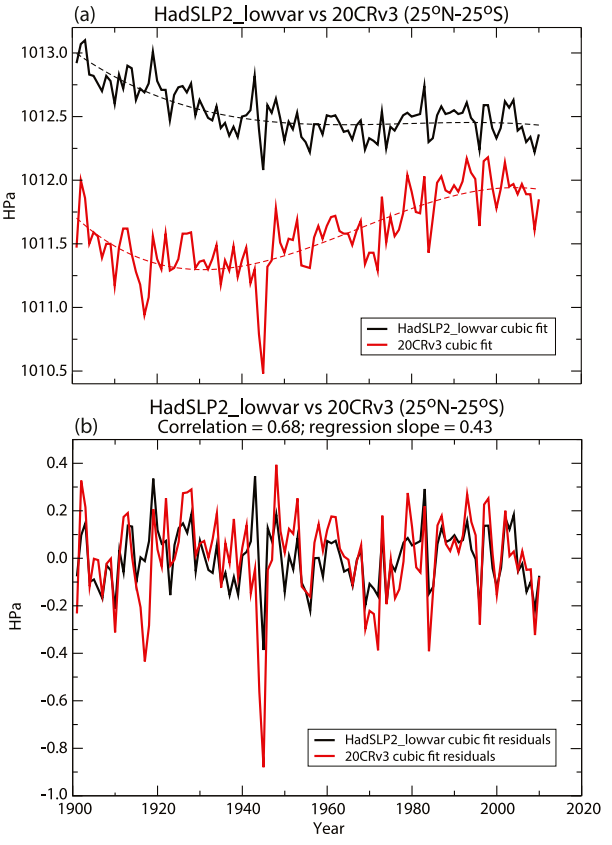

FIG. 8. (a) HadSLP2_lowvar vs 20CRv3 SLP averaged over 25°N–25°S for 1901–2010 (hPa). Dashed lines are cubic least squares fits to the data. (b) Residuals (hPa) from cubic fits in (a). Correlation between residuals was 0.68.

high-frequency variability is fairly consistent between datasets (correlation of 0.68) suggesting that the differences between the datasets are mainly of a large spatial scale (Fig. 7) and long time scale (Fig. 8).

Having documented these large-scale differences between observed data products, the question remains as to which of the observed data products is more reliable for trend analysis in the regions of strong trend differences. We speculate that the changing availability of data over time in the 20CRv3, combined with model simulation influences, could conceivably have resulted in some long-term trend issues for 20CRv3. Nevertheless, the resolution of the various discrepancies between data products is a research topic in itself, and beyond the scope of our present study. Therefore, we leave this issue unresolved for our present study; as a result, we will regard century-scale tropical trends from either dataset with relatively low confidence until this issue is better resolved. Nonetheless, we can evaluate our trend analysis results with some confidence in other regions, particularly in regions where results from the two observational data products are similar. These regions are identified through our analysis of the regional trends in 20CRv3 (Figs. 9–11, discussed below) in comparison to the earlier results of HadSLP2_lowvar (Figs. 4–6).

a. 1901–2010 trends assessment: 20CRv3

The SLP trend map assessment for the 20CRv3 dataset for 1901–2010 is shown in Fig. 9. The category assessment (Fig. 9b) indicates fairly widespread detectable increases, with some anthropogenic contribution inferred, across the Southern Hemisphere extratropics from about 25° to 40°S. This provides some support for the detectable anthropogenic increases found near this latitude band for the HadSLP2_lowvar observed dataset. However, the signal is weaker in places in the 20CRv3, which has some negative trends just poleward of 40°S in the Indian Ocean sector and near 30°S in the Atlantic sector, in contrast to the positive trends in these regions in the CMIP5 All-Forcing ensemble (Fig. 9c) and in HadSLP2_lowvar (Fig. 4a). The strong positive trends seen in the tropics (25°N–25°S) in 20CRv3 (Figs. 9a,c) are part of the strong tropical increasing trend signal just discussed in which the HadSLP2_lowvar and 20CRv3 are in strong disagreement. These tropical trends, being highly data product dependent, will not be further discussed here. Similarly, in the northern subtropical to midlatitudes (poleward of 25°N) there are more strong positive trends features for 20CRv3 (Figs. 9a,b) than for HadSLP2_lowvar (Figs. 4a,b). However, given the large differences between observed data products for these trend features, we do not have high confidence in the trends shown for either data product at this stage. Poleward of about 50°N in the relatively high latitudes, neither of the two observational data products shows substantial evidence for detectable trends over 1901–2010.

b. 1951–2010 trends assessment: 20CRv3

The 20CRv3 SLP trend maps assessment for 1951–2010 is shown in Fig. 10. The 20CRv3 data (Fig. 10a) have much stronger positive trends throughout the tropics than either the CMIP5 ensemble (Fig. 10c) or the HadSLP2_lowvar (Fig. 4a),

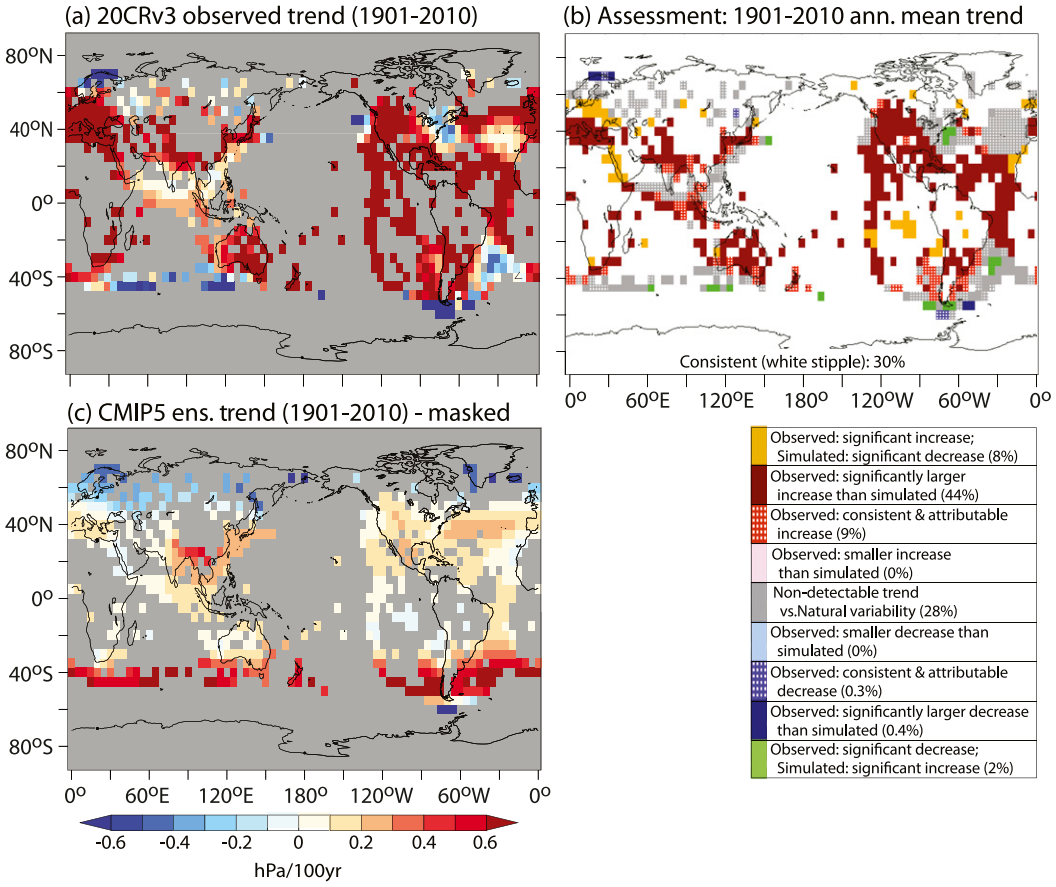

FIG. 9. As in Fig. 4, but for 20CRv3 observed trends rather than HadSLP2_lowvar (1901–2010 trend period).

extending to about 40°S in the Southern Hemisphere and over western Europe and western North America to about 50°N. However, for these low to subtropical latitude increases we will defer discussion pending further studies on the differences between 20CRv3 and HadSLP2_lowvar trends in these regions over this period. The detectable increases near 40°S in the 20CRv3 are also reflected in model ensemble trends (Fig. 10c) and provide further support for the detectable anthropogenic influence identified there in the HadSLP2_lowvar dataset (Fig. 4). As for negative SLP changes in Fig. 10, there is some resemblance between 20CRv3 and the CMIP5 models in the broad swath of negative trends spread across the high latitudes of northern Asia (65°–80°N), but only a small subset of these trends are categorized as unusual compared to Natural-Forcing runs of CMIP5 (Fig. 10b). While this feature could represent an emerging detectable negative trend, as simulated in the CMIP5 models (Fig. 10c), the results for the century scale (Fig. 9b) were much more mixed, indicating that there is not a robustly detectable century-scale SLP trend feature over these latitudes in 20CRv3.

c. 1981–2010 trends assessment: 20CRv3

The trend assessment maps for 1981–2010 for the 20CRv3 data product (Fig. 11) suggests that the regions with detectable SLP trends are mostly confined to lower (tropical and subtropical) latitudes for this period. In particular, a pronounced region of detectable increasing trends is inferred for the eastern tropical and subtropical Pacific, with some smaller regions of detectable decreasing SLP trends over the eastern North Atlantic and Africa. The eastern Pacific and eastern North Atlantic features were also detectable in the HadSLP2_lowvar assessment (Fig. 6b) suggesting that these features are robust across at least these two observational data products on this shorter time scale. Significant decreasing trends were more prevalent over the Maritime Continent and South Asia in the HadSLP2_lowvar assessment, while most grid points in these regions had nondetectable trends in the 20CRv3 data product (Fig. 11b). The HadSLP2_lowvar and 20CRv3 analyses (Fig. 6b vs Fig. 11b) agree on a general lack of detectable SLP trends in the high latitudes of either hemisphere over this time period.

5. Assessments of seasonal trends and of detectability

In this section we show trend assessment results for HadSLP2_lowvar based on individual 3-month seasons for the period 1901–2010. These can be compared to the annual mean trend assessments in Fig. 4b for HadSLP2_lowvar, or Fig. 9b for

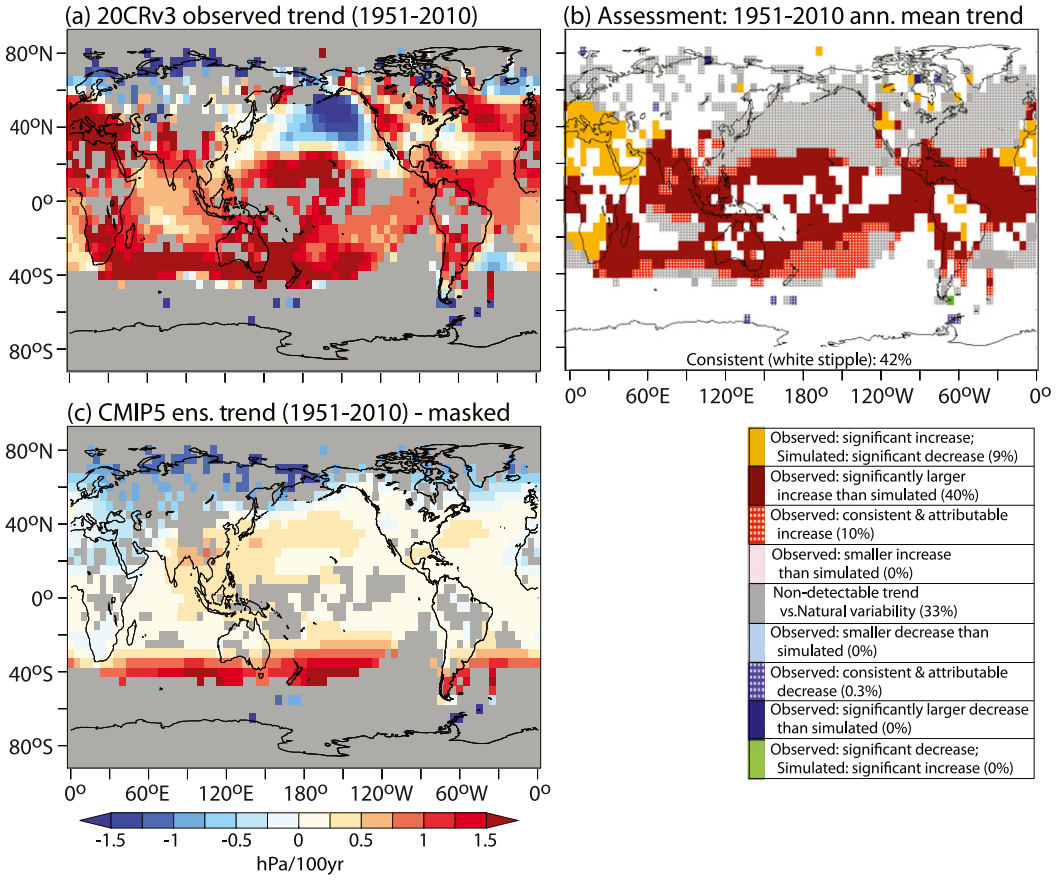

FIG. 10. As in Fig. 4, but for 20CRv3 observed trends rather than HadSLP2_lowvar, and for the period 1951–2010.

20CRv3. Seasonal versions of the trend assessments for 1951–2010 and 1981–2010 trends for HadSLP2_lowvar are contained in the supplemental material and are not discussed in detail here.

For 1901–2010 seasonal HadSLP2_lowvar trends (Fig. 12), the main result is that the detectable increasing SLP trends in the Southern Hemisphere midlatitudes in the annual mean assessment (Fig. 4b) are also present especially in the summer (DJF) season but also during MAM and JJA seasons. However, these features are only marginally detectable during Southern Hemisphere spring (SON). Elsewhere, the seasonal assessment maps indicate little evidence for detectable century-scale changes in the Northern Hemisphere extratropics during any season. The categorization results in Fig. 12c, for the July–August season, suggest some detectable increasing SLP trends over parts of central North America, southern Asia, eastern Asia, and Australia. Comparison to the annual-mean trends assessment (Fig. 4b) shows that these features were classified as detectable only for parts of eastern Asia and Australia for the HadSLP2_lowvar dataset (Fig. 4b). Further comparison to 20CRv3 annual trend results (Fig. 9b) shows that all of these increasing trend features were present to some degree in that dataset (annual mean trends). Overall, the comparison of the assessment results between HadSLP2_lowvar and 20CRv3 datasets and considering seasonal means (Figs. 4, 9, and 12) suggests robust detectable increases across the Southern Hemisphere midlatitudes, with perhaps some additional regional increases as mentioned above during the JJA season. Elsewhere and in other seasons, regional SLP trends over the twentieth century generally lack robustness across the datasets.

The issue of *detectability* of trends is an alternate way of assessing modeled trends and can be illustrated by reference to Fig. 2. Here we denote trends as having high potential for detectability if the All-Forcing ensemble mean trend is outside the 5th to 95th percentile of simulated natural variability trends over the same time period. This can be assessed using the trend distributions that were constructed for comparison to observations, but instead, comparing the All-Forcing grand ensemble trend to the Natural-Forcing trend distribution. For the example in Fig. 2 (30°–50°S trends) the All-Forcing ensemble mean trend of about 0.5 hPa $(100\,\mathrm{yr})^{-1}$ lies well outside of the Natural-Forcing average or aggregate 5th–95th percentile distributions, so that detectability is high in this region over the period 1901–2010, according to the CMIP5 models. Detectability results for various regions and time periods are summarized in map form in the supplemental material (Fig. S4). Detectability is generally higher for longer-term trends because the All-Forcing trends tend to be gradual but sustained over long periods (e.g., a century or more). We find that about 20% of our analyzed area has All-Forcing trends outside the 5th–95th percentile

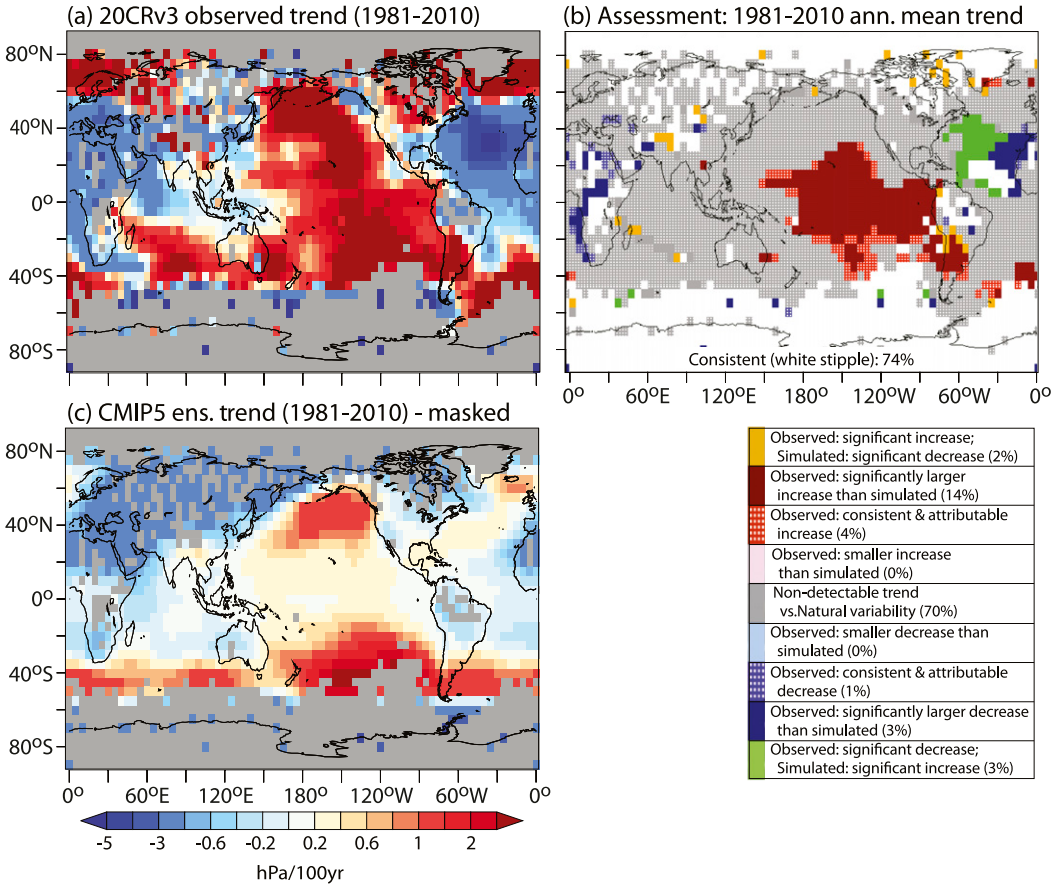

FIG. 11. As in Fig. 4, but for 20CRv3 observed trends rather than HadSLP2_lowvar, and for the period 1981–2010.

of the Natural-Forcing and Control run trend distributions for 1901–2010. These are mostly in the deep Southern Hemisphere (30°–50°S) and in and near Southeast Asia. For 1951–2010 this fraction decreases to 6%, and for 1981–2010 it decreases to 0%. That is, for trend periods as short as 30 years, we find no grid points where the All-Forcing trend is large enough to be highly unusual compared to natural variability trends on that same time scale. When we do find detectable trends in our analysis in a larger fraction of area than captured by this detectability metric, it is typically because an internal variability episode is combining with the All-Forcing trend to produce a trend that is detectable. Our analysis further confirms the usefulness of long records for detecting anthropogenic influence in the presence of internal climate variability.

6. Regional case studies

In this section, we examine time series averaged over a few selected regions to further explore the observed and modeled behavior and data quality issues related to some key large-scale trend features that emerged from the previous sections.

a. Northern Hemisphere extratropics

The trend assessment maps for HadSLP2_lowvar and 20CRv3 found some weak indication of possibly emerging anthropogenically forced negative SLP trends in higher latitudes of the Northern Hemisphere, although this feature was more evident for 1951–2010 trends than for 1901–2010 trends. To further explore this feature, time series of SLP anomalies area-averaged over the northern extratropics (45°–90°N) are shown in Fig. 13a for observations (HadSLP2_lowvar, 20CRv3) and the CMIP5 All-Forcing grand ensemble mean. The main feature in the CMIP5 multimodel mean (black curve) is a small decreasing trend, amounting to about 0.2 hPa over the twentieth century. The observed data products show small nominally positive trends over the twentieth century, and also large multidecadal variability relative to these small trends. (Internal variability is muted in the CMIP5 multimodel ensemble time series since the ensemble and multimodel averaging smooth out internal variations present in single ensemble members–which are analogous to observations.) The large variability in the observed time series, the small magnitude of trend simulated in the multimodel ensemble, and the disagreement between modeled and observed trends for the region in Fig. 13a, along with the general lack of detectable trends in the northern extratropics in the grid point based analysis for 1901–2010 (Fig. 4b), all suggest that it will be difficult to identify detectable human influence on SLP within the northern extratropics. The results further suggest that the possibly emerging negative trend signal seen for the 1951–2010 period is also not robust considering the

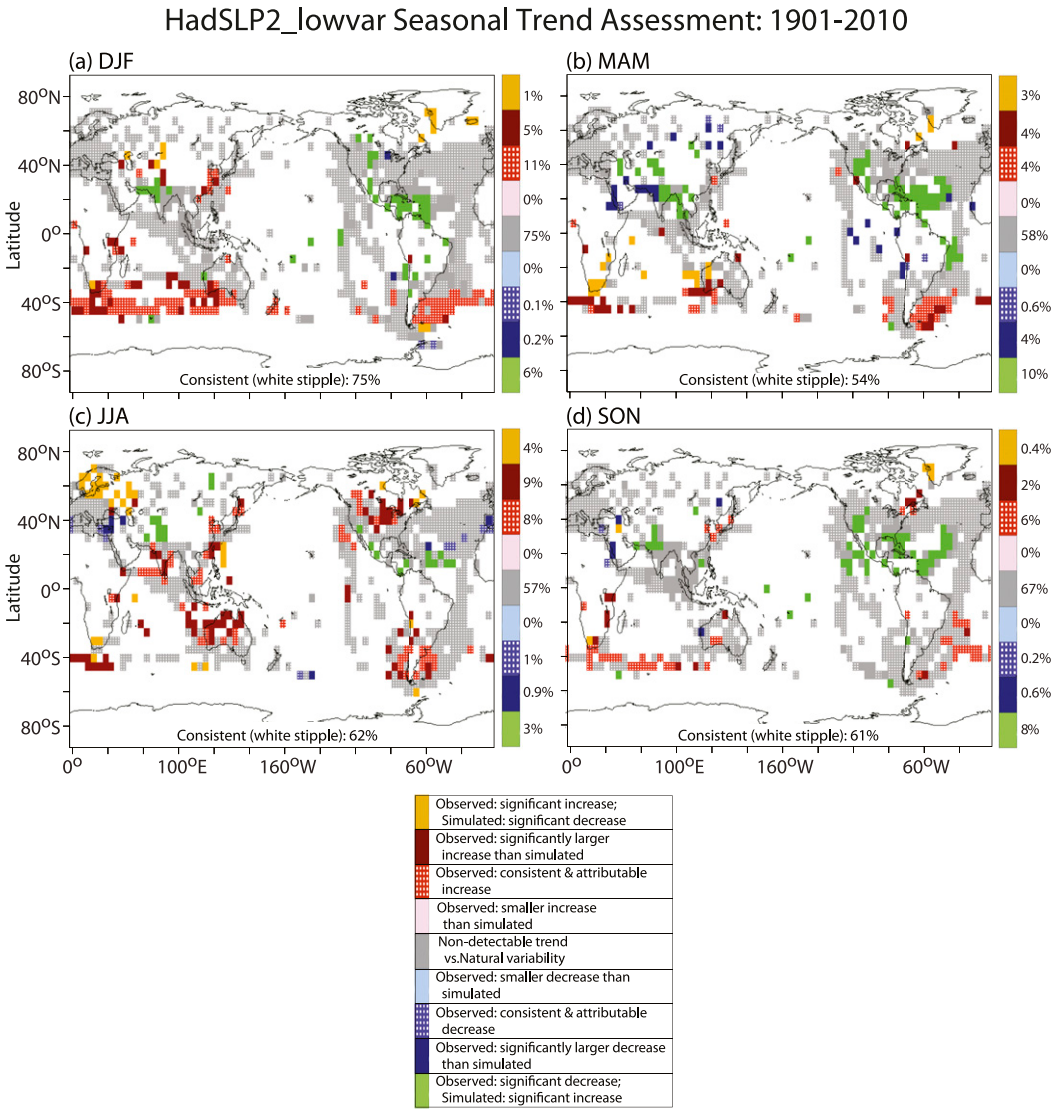

FIG. 12. As in Fig. 4b, but for 3-month seasons: (a) DJF, (b) MAM, (c) JJA, and (d) SON. Categories are described in the legend box. Percent of analyzed area for each category is listed along the color bar for each panel. White stipple regions denote where HadSLP2_lowvar observed and CMIP5 model ensemble All-Forcing trends are consistent (see text).

multidecadal variability in observations earlier in the twentieth century. We note that the early twentieth-century data are sparser and probably less reliable than the post-1950 period, so it is still possible that a negative anthropogenic trend could emerge, particularly if data quality problems are confounding trend results for this region at the century scale. There are at least a few small sets of grid points in the northern extratropics that are classified as having detectable anthropogenic influence (decreases in mean annual SLP) over 1901–2010 (Figs. 4b and 9b). For example, in north-central to northeast Russia are six grid boxes with detectable anthropogenic decreases inferred in HadSLP2_lowvar (Fig. 4b), and over extreme northern Scandinavia are several grid boxes with inferred detectable decreases in the 20CRv3 data (Fig. 9b). However, the detection results for these particular regional trend features are not robust between the two observational data products. In summary, at this stage there is little compelling evidence in the data for detectable anthropogenic influence on mean SLP over the large-scale northern extratropics.

b. Eastern equatorial Pacific region

In the eastern equatorial Pacific our observed SLP trend map analyses showed a strong positive detectable trend feature, which was particularly pronounced and detectable over 1981–2010 (Figs. 6b and 11b). CMIP5 models do not simulate as strong an increasing SLP trend (1981–2010) feature as observed in this region, compared to either the HadSLP2_lowvar or 20CRv3 dataset;

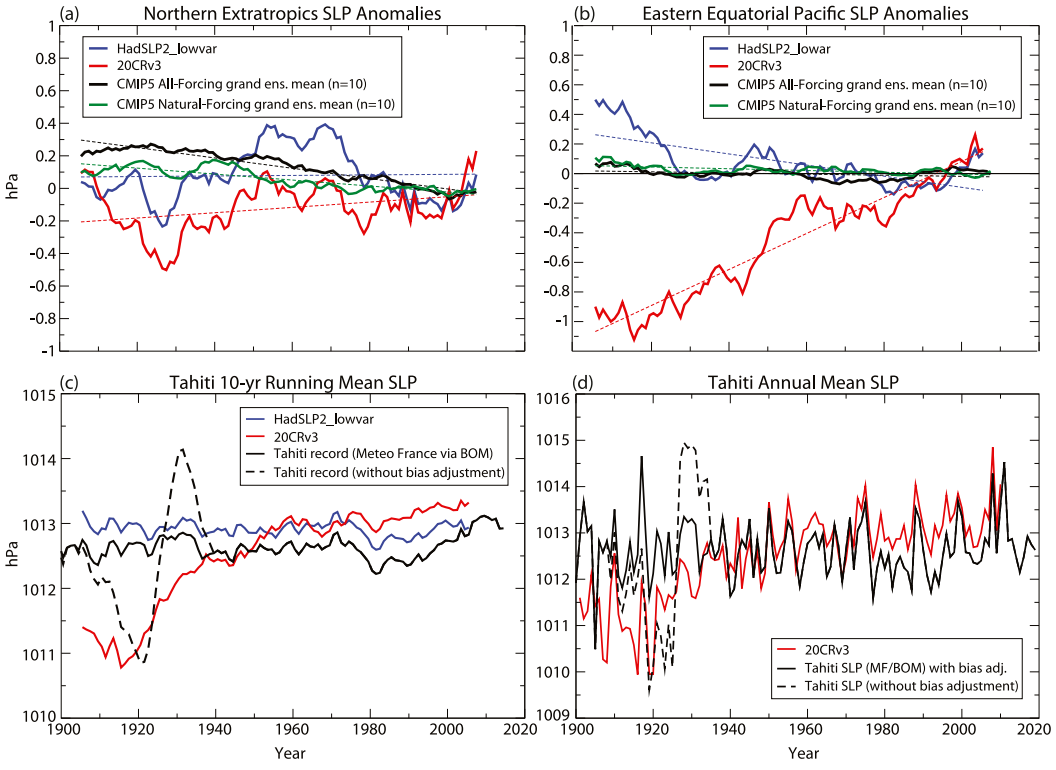

FIG. 13. Time series of 10-yr running mean SLP (hPa) averaged over (a) the northern extratropics (anomalies: 45°–90°N) and (b) the eastern equatorial Pacific (anomalies: 12.5°S–7.5°N, 150°–75°W), and (c) Tahiti vicinity SLP (HadSLP2_lowvar or 20CRv3) or Tahiti SLP reconstructions with and without bias adjustments (see text). (d) Annual mean SLP for Tahiti vicinity (20CRv3) or Tahiti SLP reconstructions with and without bias adjustment. Dashed lines in (a) and (b) depict linear trends. Figure legends identify the data source for each curve.

CMIP5 ensemble trends are particularly weak compared to the 20CRv3 data trends. To explore this feature further, Fig. 13b shows SLP time series for the region 12.5°S–7.5°N, 150°–75°W in the central and eastern equatorial Pacific. These series confirm the strong rise in SLP in the region since about 1980, particularly in the 20CRv3, with a smaller increase in HadSLP2_lowvar. Prior to 1980, however, the two observational data products show completely different long-term trend behavior, with the 20CRv3 showing a strong long-term increase of order 1 hPa per century, and the HadSLP2_lowvar showing a long-term decrease of order −0.4 hPa per century. The CMIP5 grand ensemble mean shows almost no trend over the twentieth century in this region, and only a small increase over the 1980–2010 period.

In terms of the century-scale trend differences between HadSLP2_lowvar and 20CRv3, it is of interest to consider how these could manifest in terms of measures of the Southern Oscillation index (SOI) or Walker circulation. As background, a long-term (1854–2006) weakening of the Walker circulation was reported by Vecchi et al. (2006) based on equatorial zonal wind stress. However, Hartmann et al. (2013) concluded that the observed Walker circulation weakening trend through the 1990s had been largely offset by more recent increases, calling into question whether the Walker circulation has in fact significantly weakened since the late nineteenth century (e.g., Solomon and Newman 2012). The recent rise in SLP since 1980 (Fig. 13b) is undoubtedly related to the unusually strong tropical Pacific trade winds during this period as discussed by England et al. (2014), Takahashi and Watanabe (2016), and Chung et al. (2019). These anomalies, in turn, may also be related to the failure of current climate models to simulate the trend behavior of equatorial Pacific SST anomalies since the 1950s (Seager et al. 2019).

Given the stark differences in century-scale SLP trends over the eastern equatorial Pacific between HadSLP2_lowvar and 20CRv3 datasets (Fig. 13b), we have further compared these datasets using some long SLP monthly mean time series for Tahiti and Darwin that are maintained to monitor the SOI. For example, the Bureau of Meteorology hosts online records of monthly mean SLP for Tahiti and Darwin extending from 1876 to the present (http://www.bom.gov.au/climate/current/soihtm1.shtml). We find that the long-term SLP behavior for Darwin (or Darwin vicinity for the HadSLP-lowvar and 20CRv3) is similar across the datasets (not shown). However, for Tahiti, there are large differences between the datasets (Fig. 13c, showing 10-yr running means). Apart from a uniform offset, HadSLP2_lowvar SLP in the vicinity of Tahiti has very similar low-frequency variability to the Tahiti climate records (Fig. 13c), which is not unexpected as the Tahiti data have presumably been used as input to HadSLP2_lowvar and so are not an independent data source. The 20CRv3 data, however, are very different, with

10-yr running mean SLP values in the vicinity of Tahiti that are more than 1 hPa lower than the Tahiti climate data, particularly in the early twentieth century. The Tahiti monthly mean SLP record available from the Bureau of Meteorology (BOM) site was obtained from Météo-France interregional direction for French Polynesia. For the early period 1876–1935, these data appear essentially identical to that published in Ropelewski and Jones (1987; hereafter RJ87; see their Table 1) except they include values for some periods reported as missing in RJ87. RJ87 made bias adjustments to the Tahiti dataset (their Table 2) based on comparison of Tahiti SLP with that of several other island stations in the region. The reasons for these bias adjustments were only speculated upon in RJ87, who mentioned changes in observation times and barometer elevation as possible influences. The dashed line in Fig. 13c shows the Tahiti SLP record without the RJ87 bias adjustments. Figure 13d shows a higher-frequency (annual mean) version of the SLP time series values for 20CRv3 versus the Tahiti SLP record with and without the RJ87 bias adjustments. These comparisons show that the RJ87 bias adjustments are of similar magnitude to the differences between HadSLP2_lowvar and 20CRv3 for the Tahiti region. In some cases the bias adjustments are in the opposite direction to the 20CRv3 in terms of a deviation from the unadjusted data. Overall, these results suggest some strategies by which the differences between these observational products could possibly be further reconciled (e.g., using nearby stations, metadata records of observational changes, or dynamical or statistical models), and work is underway elsewhere on examining the differences (G. Compo 2020, personal communication).

7. Summary and conclusions

In this analysis, we have assessed regional trends in historical sea level pressure (SLP) from 1901, 1951, and 1981 to 2010 in two observational products (HadSLP2_lowvar and 20CRv3). The observed trends are compared to a series of CMIP5 multimodel ensemble of simulations, including runs with no external forcing (Control runs), Natural-Forcing runs, or All-Forcing runs (natural plus anthropogenic forcings combined). The observed trends are assessed to see (i) whether they are consistent with the All-Forcing ensemble; (ii) whether they consistent with, or highly unusual compared with, the Natural Forcing and Control run distributions, which include internal variability); and (iii) whether there is evidence for detectable anthropogenic influence (defined here as regions where the observed trends are both highly unusual compared with modeled natural variability and consistent with (or at least having the same sign as) the ensemble trends produced by models that include anthropogenic as well as natural forcings.

Using a univariate detection/attribution approach, we find that the most robustly detectable anthropogenic signal is the zonal band of SLP increase over the Southern Hemisphere extratropics, seen clearly in the 1901–2010 and 1951–2010 trends. While a lack of global data coverage over the period in observations precludes a complete description of this feature, in models it is associated with a negative trend of SLP in higher southern latitudes, with a nodal line in the SLP trends around 60°S. Previous studies indicate that this long-term circulation change is induced in models by greenhouse gas increases and tropospheric and stratospheric ozone changes (Gillett et al. 2013). The positive SLP trend feature, present over roughly 30°–50°S, is fairly well simulated by the CMIP5 All-Forcing ensemble. Our results support previous findings of detectable anthropogenic influence in this region.

A second finding from our study is that there is very limited evidence for detectable anthropogenic influence on SLP trends in the Northern Hemisphere middle and high latitudes. This is at least qualitatively consistent with the conclusions of IPCC AR5, which for example did not conclude that detectable anthropogenic influence could be identified specifically in the region (e.g., in such indices as the northern annular mode).

Within the lower latitudes (25°–25°S and some surrounding regions), we found that trends from HadSLP2_lowvar and 20CRv3 were remarkably different, particularly on the century scale. For example, a comparison of area-averaged time series for 25°N–25°S shows a century-scale increasing trend in 20CRv3, with a century-scale decreasing trend in HadSLP2_lowvar. These differences are large and impact any detection/attribution and consistency conclusions for these regions. The tropical Pacific is a particular region with large differences, as exemplified by analyses of SLP time series for the Tahiti region and for the central and eastern equatorial Pacific. Therefore, at this stage, we can make no firm conclusions about detection/attribution or consistency of long-term (1901–2010 or 1951–2010) SLP trends in these low-latitude regions, pending further investigation to resolve key differences between the observational data products we have examined.

Thus, from a global perspective, while we find a fairly large fractional area where trends based on the "average characteristics" of the CMIP5 All-Forcing ensemble are classified as being inconsistent with observed trends, any firm conclusion on this inconsistency at the global scale is premature at this stage, given the aforementioned observational data issues. We note that at least the zonal band of SLP increase over the Southern Hemisphere extratropics is relatively robust across observational data products, with relatively robust detection and attribution in part to anthropogenic forcing inferred from our analysis. This supports earlier detection/attribution findings on this and related circulation features.

It is generally much more difficult to identify detectable anthropogenic signals in the 1981–2010 observed regional trends than for trends over 1901–2010 or 1951–2010. This may be because it is much more difficult to identify detectable anthropogenic influence over such a short period as 1981–2010 due to large internal variability on such time scales, and the low potential detectability of gradual long-term trends over short periods for SLP (e.g., Fig. S4). Further, our analysis shows that the observed SLP trend patterns over 1981–2010 are quite different from the simulated externally forced trend pattern over this same time period. Our results suggest that internal variability is likely the dominant driver of most of the 1981–2010 regional trend features.

As a further example of this, a prominent regional SLP trend feature over 1981–2010 is the detectable increase in SLP in the central equatorial Pacific seen in the 1981–2010 trend

assessment maps. This increase is poorly modeled because the positive trend is significantly larger in observations than in the CMIP5 ensemble trend distribution, even accounting for model simulated internal variability. Thus in our assessment, the observed 1981–2010 trend in the eastern tropical Pacific is highly unusual compared to either the CMIP5 average simulated natural variability or to the average modeled All-Forcing response (i.e., it is difficult to reconcile with the CMIP5 All-Forcing multimodel ensemble).

Overall, our analysis strongly suggests that some detectable anthropogenic influences are present in regional SLP trends extending over 1901–2010 or 1951–2010, especially in the Southern Hemisphere extratropics. However, a number of multidecadal SLP trend features remain poorly simulated by models, such as in the equatorial Pacific in recent decades. Little evidence is found for detectable anthropogenic influence in the northern extratropics. Finally, important differences were found in long-term trend behavior between HadSLP2_lowvar and 20CRv3 datasets, which require further investigation.

Acknowledgments. We thank Peter Phillipps and Liwei Jia for providing constructive comments on a draft manuscript. We thank R. Allan, T. Ansell and the UKMO Hadley Centre for developing and providing the HadSLP2 dataset online; Gil Compo and Laura Slivinski for developing and providing the 20CRv3 dataset and for helpful discussions on this dataset; and the CMIP5 modeling groups for contributing simulations to the CMIP5 online archive.

REFERENCES

Allan, R., and T. Ansell, 2006: A new globally complete monthly historical gridded mean sea level pressure dataset (HadSLP2): 1850–2004. *J. Climate*, **19**, 5816–5842, https://doi.org/10.1175/JCLI3937.1.

Allen, R. J., and M. Kovilakam, 2017: The role of natural climate variability in recent tropical expansion. *J. Climate*, **30**, 6329–6350, https://doi.org/10.1175/JCLI-D-16-0735.1.

Bhend, J., and P. Whetton, 2013: Consistency of simulated and observed regional changes in temperature, sea level pressure, and precipitation. *Climatic Change*, **118**, 799–810, https://doi.org/10.1007/s10584-012-0691-2.

Bindoff, N. L., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952.

Blackport, R., J. A. Screen, K. van der Wiel, and R. Bintanja, 2019: Minimal influence of reduced Arctic sea ice on coincident cold winters in mid-latitudes. *Nat. Climate Change*, **9**, 697–704, https://doi.org/10.1038/s41558-019-0551-4.

Chung, E.-S., A. Timmermann, B. J. Soden, K.-J. Ha, L. Shi, and V. O. John, 2019: Reconciling opposing Walker circulation trends in observations and model projections. *Nat. Climate Change*, **9**, 405–412, https://doi.org/10.1038/s41558-019-0446-4.

Cohen, J., and Coauthors, 2014: Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, **7**, 627–637, https://doi.org/10.1038/ngeo2234.

——, K. Pfeiffer, and J. A. Francis, 2018: Warm Arctic episodes linked with increased frequency of extreme winter weather in the United States. *Nat. Commun.*, **9**, 869, https://doi.org/10.1038/s41467-018-02992-9.

Coumou, D., G. Di Capua, S. Vavrus, L. Wang, and S. Wang, 2018: The influence of Arctic amplification on mid-latitude summer circulation. *Nat. Commun.*, **9**, 2959, https://doi.org/10.1038/S41467-018-05256-8.

England, M. H., and Coauthors, 2014: Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus. *Nat. Climate Change*, **4**, 222–227, https://doi.org/10.1038/nclimate2106.

Francis, J. A., and S. J. Vavrus, 2015: Evidence for a wavier jet stream in response to rapid Arctic warming. *Environ. Res. Lett.*, **10**, 014005, https://doi.org/10.1088/1748-9326/10/1/014005.

Gillett, N. P., and P. A. Stott, 2009: Attribution of anthropogenic influence on seasonal sea level pressure. *Geophys. Res. Lett.*, **36**, L23709, https://doi.org/10.1029/2009GL041269.

——, F. W. Zwiers, A. J. Weaver, and P. A. Stott, 2003: Detection of human influence on sea-level pressure. *Nature*, **422**, 292–294, https://doi.org/10.1038/nature01487.

——, R. J. Allan, and T. J. Ansell, 2005: Detection of external influence on sea level pressure with a multi-model ensemble. *Geophys. Res. Lett.*, **32**, L19714, https://doi.org/10.1029/2005GL023640.

——, J. C. Fyfe, and D. E. Parker, 2013: Attribution of observed sea level pressure trends to greenhouse gas, aerosol, and ozone changes. *Geophys. Res. Lett.*, **40**, 2302–2306, https://doi.org/10.1002/grl.50500.

Grise, K. M., S. M. Davis, P. W. Staten, and O. Adam, 2018: Regional and seasonal characteristics of the recent expansion of the tropics. *J. Climate*, **31**, 6839–6856, https://doi.org/10.1175/JCLI-D-18-0060.1.

Hartmann, D. L., and Coauthors, 2013: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 159–254.

Hegerl, G. C., and Coauthors, 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 663–745.

Knutson, T. R., and J. J. Ploshay, 2016: Detection of anthropogenic influence on a summertime heat stress index. *Climatic Change*, **138**, 25–39, https://doi.org/10.1007/s10584-016-1708-z.

——, and F. Zeng, 2018: Model assessment of observed precipitation trends over land regions: Detectable human influences and possible low bias in model trends. *J. Climate*, **31**, 4617–4637, https://doi.org/10.1175/JCLI-D-17-0672.1.

——, ——, and A. T. Wittenberg, 2013: Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth century simulations. *J. Climate*, **26**, 8709–8743, https://doi.org/10.1175/JCLI-D-12-00567.1.

——, and Coauthors, 2019: Tropical cyclones and climate change assessment: Part I: Detection and attribution. *Bull. Amer. Meteor. Soc.*, **100**, 1987–2007, https://doi.org/10.1175/BAMS-D-18-0189.1.

Masson-Delmotte, V., and Coauthors, 2013: Information from paleoclimate archives. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 383–464.

Ropelewski, C. F., and P. D. Jones, 1987: An extension of the Tahiti-Darwin Southern Oscillation Index. *Mon. Wea. Rev.*, **115**, 2161–2165, https://doi.org/10.1175/1520-0493(1987)115<2161:AEOTTS>2.0.CO;2.

Seager, R., M. Cane, N. Henderson, D.-E. Lee, R. Abernathy, and H. Zhang, 2019: Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse

gases. *Nat. Climate Change*, **9**, 517–522, https://doi.org/10.1038/s41558-019-0505-x.

Slivinski, L. C., and Coauthors, 2019: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quart. J. Roy. Meteor. Soc.*, **145**, 2876–2908, https://doi.org/10.1002/qj.3598.

Solomon, A., and M. Newman, 2012: Reconciling disparate twentieth-century Indo-Pacific Ocean temperature trends in the instrumental record. *Nat. Climate Change*, **2**, 691–699, https://doi.org/10.1038/nclimate1591.

Sun, C., and Coauthors, 2019: Spring Aleutian low weakening and surface cooling trend in northwest North America during recent decades. *J. Geophys. Res. Atoms.*, **124**, 12 078–12 092, https://doi.org/10.1029/2019JD031405.

Swart, N. C., and J. C. Fyfe, 2012: Observed and simulated changes in the Southern Hemisphere surface westerly wind-stress. *Geophys. Res. Lett.*, **39**, L16711, https://doi.org/10.1029/2012GL052810.

Takahashi, C., and M. Watanabe, 2016: Pacific trade winds accelerate by aerosol forcing over the past two decades. *Nat. Climate Change*, **6**, 768–772, https://doi.org/10.1038/nclimate2996.

Thompson, D. W. J., S. Solomon, P. J. Kushner, M. H. England, K. M. Grise, and D. J. Karoly, 2011: Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nat. Geosci.*, **4**, 741–749, https://doi.org/10.1038/ngeo1296.

Varino, F., P. Arbogast, B. Joly, G. Riviere, M.-L. Fandeur, H. Bovy, and J.-B. Granier, 2019: Northern Hemisphere extratropical winter cyclones variability over the 20th century derived from ERA-20C reanalysis. *Climate Dyn.*, **52**, 1027–1048, https://doi.org/10.1007/s00382-018-4176-5.

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

Vose, R. S., D. R. Easterling, K. E. Kunkel, A. N. LeGrande, and M. F. Wehner, 2017: Temperature changes in the United States. *Climate Science Special Report: Fourth National Climate Assessment*, D. J. Wuebbles et al. Eds., Vol. I, U.S. Global Change Research Program, 185–206, https://doi.org/10.7930/J0N29V45

Wunsch, C., 1999: The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. *Bull. Amer. Meteor. Soc.*, **80**, 245–256, https://doi.org/10.1175/1520-0477(1999)080<0245:TIOSCR>2.0.CO;2.