11-2017

Regional Sea Level Variability and Trends, 1960-2007: A Comparison of Sea Level Reconstructions and Ocean Syntheses

M. Carson

A. Köhl

D. Stammer

B. Meyssignac

J. Church

See next page for additional authors

Original Publication Citation

Carson, M., Kohl, A., Stammer, D., Meyssignac, B., Church, J., Schroter, J., . . . Hamlington, B. (2017). Regional sea level variability and trends, 1960-2007: A comparison of sea level reconstructions and ocean syntheses. *Journal of Geophysical Research: Oceans, 122*(11), 9068-9091. doi:10.1002/2017jc012992

Authors
M. Carson, A. Köhl, D. Stammer, B. Meyssignac, J. Church, J. Schröter, M. Wenzel, and B. Hamlington

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2017JC012992

Key Points:
- Regional sea level anomalies reconstructed from historical data compared
- Substantial disagreement between reconstructions before satellite altimetry
- Historical regional sea level variability and trends remain elusive

Supporting Information:
- Supporting Information S1

Correspondence to:
M. Carson,
mark.carson@uni-hamburg.de

Citation:
Carson, M., Köhl, A., Stammer, D., Meyssignac, B., Church, J., Schröter, J., . . . Hamlington, B. (2017). Regional sea level variability and trends, 1960–2007: A comparison of sea level reconstructions and ocean syntheses. *Journal of Geophysical Research: Oceans*, *122*, 9068–9091. https://doi.org/10.1002/2017JC012992

Received 13 APR 2017
Accepted 17 OCT 2017
Accepted article online 25 OCT 2017
Published online 22 NOV 2017

Regional Sea Level Variability and Trends, 1960–2007: A Comparison of Sea Level Reconstructions and Ocean Syntheses

M. Carson[1], A. Köhl[1], D. Stammer[1], B. Meyssignac[2], J. Church[3], J. Schröter[4], M. Wenzel[4], and B. Hamlington[5]

[1]Center für Erdsystemforschung und Nachhaltigkeit, University of Hamburg, Hamburg, Germany, [2]LEGOS, CNES, Toulouse, France, [3]Climate Change Research Center, University of New South Wales, Sydney, NSW, Australia, [4]Alfred Wegener Institute, Bremerhaven, Germany, [5]Ocean, Earth and Atmospheric Sciences, Old Dominion University, Norfolk, VA, USA

Abstract Several existing statistical and dynamical reconstructions of past regional sea level variability and trends are compared with each other and with tide gauges over the 48 year period 1960–2007, partially predating the satellite altimetry era. Evaluated statistical reconstructions were built from tide-gauge data (TGR), and dynamical reconstructions from ocean data assimilation (ODA) approaches. Although most of the TGRs yield global-mean time series of sea level with trends deviating within only ± 0.1 mm yr^{-1}, the spatial anomalies of the trends deviate substantially between the reconstructions over the period predating altimetry. In contrast, TGRs match observed regional trend patterns fairly well during the satellite altimetry era. TGRs match tide-gauge data better than ODA results; however, they exhibit less variability in the open ocean compared to altimetric data. Over the prealtimetry period, all reconstructed regional sea level trend patterns deviate substantially from each other. In terms of detrended correlations in this earlier period, the reconstructions match tide gauges, and each other, much better in the Pacific than in the Atlantic. An ensemble of all TGR and ODA estimates provides some improvements in correlations and trends to both tide gauges and altimetry. Nevertheless, a lack of independent open ocean sea surface height data predating altimetry makes impossible the validation of the ensemble for prealtimetry open ocean sea level trends and variability. Estimating regional sea level changes prior to altimetry therefore remains an unsolved challenge.

1. Introduction

Understanding the causal relationship between observed sea level changes and underlying processes is an important aspect of ongoing sea level research and a prerequisite for identifying anthropogenic influences on climate. Today, altimetric satellite measurements of sea surface height (SSH) are used to study sea level variability, jointly with steric height information available from Argo in situ temperature and salinity (density) measurements (e.g., Roemmich & Owens, 2000), and ocean mass changes inferred from gravimetry (e.g., GRACE data; Tapley et al., 2004). In conjunction with output from numerical ocean models (e.g., Bromirski et al., 2011; Han et al., 2010), these observations have revealed a complex picture of global and regional sea level variations and have considerably improved our understanding of present-day sea level changes in terms of ocean variability and underlying changes in ocean mass, density, and circulation. Using this information to improve coupled climate models will produce more accurate projections of future sea level rise.

Unfortunately, required high-quality observations of regional sea level changes, continuous in time and with quasi-global coverage, exist only since the advent of precise satellite altimetry in 1992 (Fu & Cazenave, 2001). Obtaining observational information on spatial sea level variations prior to the altimetry era therefore remains a challenge due to the lack of sufficient historical in situ or satellite ocean observations. As a result, over most of the ocean, historical regional sea level change estimates do not exist, this holds especially for the Southern Hemisphere. Over historical periods even as short as the last 40–50 years, regional SSH variations can only be inferred by extending the sparsely available and spatially inhomogeneous tide-gauge data, or by estimating dynamical and thermosteric SSH in ocean reanalyses constrained by scattered temperature and very limited salinity data.

To address this problem, regional sea level reconstructions have been built using statistical techniques, which combine long tide gauge records of limited spatial coverage with shorter time scale sea level variability

patterns of near-global coverage, available either from satellite altimetry or from numerical ocean models. Using these techniques, a number of sea level reconstructions (Church et al., 2004; Meyssignac et al., 2012a; Ray & Douglas, 2011) have been constructed based on EOF methods such as reduced space optimal interpolation (RSOI; Kaplan et al., 1997). Wenzel and Schröter (2010, 2014) complement these studies by using a neural network for filling data gaps, and an interpolation method which involves finding mapping weights by fitting the tide-gauge data to the EOFs to approximate the altimetric patterns, instead of the reverse.

Ocean models, driven by atmospheric reanalyses and constrained by the few historic ocean observations available, are a valuable alternative to those statistical approaches and will add useful additional information to observation-based reconstructions of past sea level changes on regional scales. Several existing ocean reanalysis products can now be used to estimate the time varying ocean state over the last 60 years, the results of which can be used to study spatial sea level variations and underlying causes (Balmaseda et al., 2013; Carton et al., 2005; Carton & Giese, 2008; Lombard et al., 2009; Song & Colberg, 2011; Wunsch & Heimbach, 2007). Köhl and Stammer (2008) and Köhl (2015) applied the ECCO state estimation techniques to estimate regional SSH changes over the last 50 years. Stammer et al. (2010) and Storto et al. (2015) compared SSH trends estimated by various ocean reanalysis approaches and highlighted differences in SSH trends from those results, especially for halosteric SSH trends. Chepurin et al. (2014) compared SSH data from ocean data reanalyses to tide gauges, similar to our approach here, but our goal is to additionally compare these products to the tide gauge reconstructions. For a review of ocean data assimilation in support of climate research, see Stammer et al. (2013).

From the above observational and reanalysis based studies, it is reasonably concluded that estimates of spatial SSH variability and trend patterns are significantly different over the past 50 years from those observed over the recent altimetry era, have a much lower magnitude, and in some regions even have opposite trends. This suggests that regional trends estimated from altimetry do not represent long-term trends due to anthropogenic forcing of the ocean, but instead very likely represent primarily low-frequency internal dynamical modes (Meyssignac et al., 2012b; White et al., 2005; Zhang & Church, 2012). The large amplitude of low-frequency, multidecadal dynamical ocean variability can make the process of estimating expected anthropogenic regional sea level changes from sparse observations complicated (e.g., Carson et al., 2015; Palanisamy et al., 2015).

The purpose of this paper is to revisit the question of how confident we are in existing reconstructions of regional sea level changes relative to the global mean. The comparison is performed among reconstructions available over their common period of 1960–2007, which explores the existence of consistent explanations of open ocean natural and anthropogenic variability and trend patterns over the recent past. Additionally, our goal is to compare available ocean reanalyses—built from models using data as constraints through data assimilation—with reconstructions generated through statistical assumptions. In this context, comparing only regional sea level information avoids a major difficulty arising from a principle difference between tide-gauge-based reconstructions and ocean data assimilations, namely the contribution from mass additions to the ocean through land ice melt, because at present this term is largely a global value with only small regional differences over most of the ocean, though there are places where effects on trends may be more noticeable (Slangen et al., 2014). The study period is split into two parts: the prealtimetry period of 1960–1992 and the altimetry period of 1993–2007. However, we also explore to what extent an ensemble mean time series or multimethod approach can lead to more accurate estimates over the entire period than individual estimates. Understanding the differences and intrinsic uncertainties in existing estimates of interdecadal regional sea level change, many of which are only model based, is an important step toward improving estimates of natural variability, which can help separating them from anthropogenic signals.

The structure of the paper is as follows: in the following section, the data products under investigation are briefly described with important differences highlighted, along with the comparison statistics used. Section 3 is broken into subsections which contain the robust features and differences grouped by characteristics, such as long-term trends, variability, and common spatial patterns. Section 4 contains concluding comments and recommendations.

2. Data and Methods

In our analysis, we use several statistical tide-gauge-based reconstructions (TGR) along with dynamical ocean reanalysis (also called ocean data assimilation, ODA) products. We selected products which had long

Table 1
Tide Gauge Reconstruction and Ocean Assimilation Products

Product	Method	Domain (lat. and time)	Original resolution
Tide gauge reconstructions			
AWI[a]	ANW + EOF	−90:90, 1901–2007	1 × 1, mth
LEGOS[b]	EOF	−50:70, 1960–2012	1 × 1, mth
CSIRO[c]	EOF	−65:65, 1950–2012	1 × 1, mth
CCAR[d]	Cyclostationary EOF	−75:89, 1950–2008	0.5 × 0.5, week
Ocean data assimilations			
GECCO2[e]	4DVar: SLA,T,S,MDT,SST	−90:90, 1960–2013	1 × 1, mth
ORA-S4[f]	3DVar: SLA,T,S,SST	−90:90, 1958–2014	1 × 1, mth
SODA 2.1.6[g]	OI: T,S,SST	−90:90, 1958–2008	0.5 × 0.5, mth

[a]Wenzel and Schröter (2010, 2014); ANW is Artificial Neural Network, which was used in gap-filling.
[b]Meyssignac et al. (2012a).
[c]Church et al. (2004).
[d]Hamlington et al. (2011).
[e]Köhl (2015).
[f]Balmaseda et al. (2013).
[g]Carton and Giese (2008).

enough coverage in time and also contained a good mixture of varying techniques. For a summary of all reconstructions, see Table 1.

2.1. Tide-Gauge-Based Reconstructions

Four TGRs are compared to tide gauges, satellite altimetry, and each other to explore similarities, strengths, and weaknesses. These reconstructions attempt to fill in the missing regional SSH data in the open ocean by various means which include fitting tide gauges to a series of empirical orthogonal functions (EOFs) which represent the patterns of open ocean variability (RSOI; Kaplan et al., 1997). These spatiotemporal patterns of SSH variability are mostly derived from satellite altimetry, which gives us the only historical observations of nearly synoptic regional SSH over most of the ocean.

The LEGOS products (an update of the product in Meyssignac et al., 2012a) use various input fields to estimate regional SSH variability patterns, including altimetry and ocean data assimilation products. These products are the only reconstruction in this study to use input fields not derived solely from altimetry, and also from a fixed number of tide gauges over the entire 1960–2009 period (updated to 2012; i.e., no additional tide gauges providing data only for some period, e.g., the 1990s, were used; only gauges with data spanning the whole period). For the LEGOS reconstructions based on patterns from data assimilation products, the loading patterns were calculated from the entire period covered by the product (at least as far back as 1960), in contrast to the reconstruction based on altimetry data. In total, there are five LEGOS-produced data sets, which all performed similarly in statistical tests. Therefore, we averaged them into one data product to compare to the other reconstructions, which reduces the number of comparisons and allows the study of one "ensemble mean." Some concluding remarks regarding the performance of the averaged LEGOS reconstruction are provided in section 4. One preliminary comment which will be shown relevant to the comparisons below: the LEGOS reconstructions are based on EOF loading patterns taken on basin-wide data separately; namely, the Atlantic basin is resolved separately from a single Indo-Pacific basin.

The CSIRO product is an update of the original product described in Church et al. (2004), with the new version extended to 2012. The base loading patterns are a subset of global EOFs from altimetry needed to explain the majority of SSH variability, and then are fit to a selected set of tide gauges in a least squares sense to provide the reconstructed SSH fields. This is almost the same method as used for LEGOS, but with global EOFs. The reconstruction from CCAR uses a similar approach, but instead of using regular EOFs derived from altimetry, cyclostationary EOFs are employed (Hamlington et al., 2011, 2012). These are EOFs which can vary with a set periodicity, and can therefore represent, e.g., the annual cycle in a single mode. A somewhat reversed approach is used in the AWI reconstruction where, instead of fitting the altimetry-derived EOFs to the tide gauges to obtain EOF weightings for reconstructing historical SSH, the tide gauges

were fit to the EOFs in order to obtain weights for reconstructing the altimetry data (Wenzel & Schröter, 2010, 2014) over the altimetric era. The resulting weights are then assumed to apply to the entire period of the historical SSH reconstruction.

The different groups constructing TGRs included differing sets of tide gauges data in their reconstructions. These data sets were based on the planned temporal coverage (AWI and CSIRO reconstruct further back in time than LEGOS and CCAR) as well as for the selection and rejection criteria. The LEGOS reconstructions used 91 tide gauges which covered the majority of the reconstructed period (1960–2012), AWI used 178, CSIRO used over 200 for nearly all of the common period in this paper, and CCAR used 409 which, for these latter three reconstructions, the number of available tide gauges differ as a function of time. All reconstruction analysis methods employed gap-filling for the tide-gauge data to some degree. All but CCAR used monthly tide-gauge data and produced monthly reconstructions, whereas CCAR used weekly data frequency for both the input data and the reconstruction. We averaged the data into annual means for our intercomparison. All TGR products were already gridded to a common $1° \times 1°$ grid by their respective groups. The common spatial domain between all products runs from 50°S to 65°N; this is the region over which all analyses here are carried out. For the time domain, the AWI reconstruction is the shortest by ending in 2007; CCAR ends in 2009. Therefore, all comparisons are evaluated over their common time frame 1960–2007. Data processing steps included correcting tide gauges for the inverse barometer effect (Wunsch & Stammer,1997) and for glacial isostatic adjustment (GIA; Church et al., 2004; Peltier, 2004) for all reconstructions. CSIRO used a slightly different ice history than the other reconstructions and the Mitrovica viscosity model, but for the purposes of this intercomparison, no noticeable differences could be detected from simply using the Peltier (2004) GIA data. For further information on all of these data processing choices, please consult the relevant paper for the reconstruction, listed in Table 1.

The GMSL time series for the TGRs, and their global-mean trends, are shown in Figure 1. For three of the four products, the estimated 1960–2007 trends are almost identical and are within 0.1 mm yr^{-1} of each other. This could be interpreted as evidence that a reasonably robust global estimate of historical sea level change is provided by these various analyses. However, these three TGRs use a similar method of capturing the global-mean signal, which is the additional of an "EOF0" that is "flat," i.e., normalized the same value over the whole ocean. This method that was used early on by Church et al.

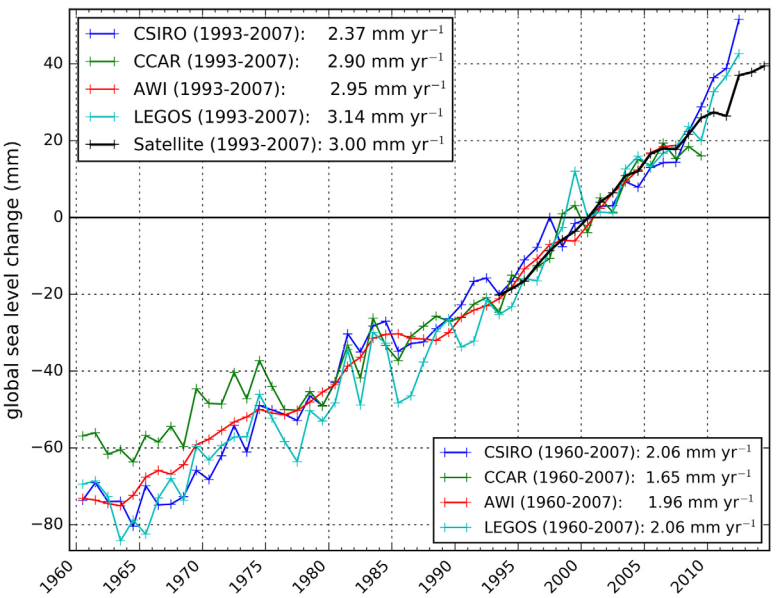

Figure 1. Global-mean sea level. Global-mean sea surface height time series from the four tide gauge reconstructions. The three other than CCAR correspond very well in overall trend, though some details in variability differ. CCAR has a markedly weaker trend. Satellite altimetry data are from the AVISO Ssalto/Duacs multimission gridded data set of sea level anomalies.

(2004) and was discussed by Calafat et al. (2014), happens to yield values close to the simple, area-weighted (on a 1 × 1 grid) tide gauge mean (not shown, but cf., Church & White, 2011, Figure 6). Tests on a reconstructed data set in Hay et al. (2015) found that the methods of Church et al. (2004) might lead to biased results, and the GMSL in the Hay et al. (2015) study is similar in trend value to Dangendorf et al. (2017). The CCAR product estimates the GMSL using a more complicated analysis (Hamlington et al., 2012). We also note that the amount of global decadal variability in the various products is not the same, though it is generally small compared to the trend. Generally, reconstructions have difficulty reproducing interannual variability which leads to differences in the GMSL time series (Calafat et al., 2014). Although the trends compare well to altimetry, and each other throughout the altimetry period, there are still biases in satellite altimetry data which are being evaluated that can affect the global-mean trend (e.g., Ablain et al., 2014; Watson et al., 2015), and this may also have some impact on TGR global or regional trends.

For sections 3.2 and later, these GMSL series were removed from the TGRs in each grid box, so that the remainder is the regional SSH anomaly relative to the global mean.

2.2. Ocean Data Assimilations

As a point of comparison for the TGRs, regional SSH anomalies from a set of three ODAs are also analyzed. GECCO2 (Köhl, 2015) and ORA-S4 (Balmaseda et al., 2013) both assimilate altimetry data (sea level anomalies, SLA; Table 1), whereas SODA v2.1.6 does not (Carton & Giese, 2008).

Because of inherent model approximations, global-mean sea level changes cannot be studied from these results (Greatbatch, 1994; Griffies & Greatbatch, 2012), and assimilation products only diagnose SSH anomalies, except for ORA-S4, which includes the altimeter GMSL. Because the other two ODAs still contain some nonzero global residual, the global mean of all three assimilation products was removed from each data set at each time step, just as for the TGRs, with the exception of section 3.1.

All data were then averaged into annual means and on the same 1° × 1° grid as the other products. The common spatial domain for all products is 50°S–65°N. Although the ODAs have more coverage than this, their domains were restricted to where the TGRs also have data.

2.3. Data Concerns

It is important to note that sea level reconstructions and models do not necessarily represent the same sea level. In particular, sea level reconstructions include land ice melt and additional ocean mass variations such as land water storage in their sea level estimates, while most process-oriented models do not, although models with data assimilation are constrained by observations. Ocean models essentially reflect the regional variability of steric effects plus mass redistribution, although internal variability not adequately captured by lower resolution models is likely missing (e.g., Serazin et al., 2016), while tide gauge reconstructions also include any processes which affect the tide gauge measurements, including local vertical land motion which may not be accounted for in their applied GIA corrections (from ICE-5G/VM2, except for CSIRO; Church et al., 2004; Peltier, 2004), tectonic shifts (Ballu et al., 2011), plus mass redistribution and changes due to ongoing land ice mass loss. The removal of the global mean from all products, and from the tide gauges used for comparison (see below), resolves a large part of ice mass loss and land water storage problems.

In order to compare the quality of the tide gauge reconstructions to real ocean SSH data, the only data set which can be used in the prealtimetry era is from the tide gauges themselves. In many locations, tide-gauge data also do not reflect the same SSH as in the open ocean measured by satellite altimetry due to meteorological forcing (e.g., Sturges & Douglas, 2011; Wunsch & Stammer, 1997), and various sources of vertical land motion (Ballu et al., 2011; Cahoon, 2015; Kolker et al., 2011; Peltier, 2004) at the gauge's location. It was clear in the reconstruction process that vertical land motion needed to be corrected. Commonly used GIA data sets were applied to the tide-gauge data before the reconstructions were computed. However, recent studies have pointed to the lack of sufficiently accurate estimates of vertical land motion at tide gauge locations for the purposes of computing GMSL (Dangendorf et al., 2017; Hay et al., 2015; King et al., 2012). While we focus on regional trends and variability in this paper, residual vertical land motion may also be the cause of problems in the TGRs estimating regional trends (see section 3.2). Some further comments on this concern are included in section 4.

2.4. Comparison Methods

During this study, we used tide-gauge data to analyze the quality of existing TGRs and ODAs reconstructions—together referred to as the "products"—by doing two sets of comparisons: (1) regional trends (with the global mean removed) and (2) calculating correlations and root-mean-square deviations of colocated data with respect to tide gauges and satellite altimetry. Trend comparisons to tide-gauge data are done only for regional sea level, thus some estimate of GMSL has to be removed from them. Since there are four different estimates of GMSL from the four TGRs, and only one ODA has a GMSL estimate (ORA-S4), the GMSL removed from the tide gauges is the mean of the three TGR estimates that lie closest together (AWI, CSIRO, and the LEGOS mean). These were chosen since this choice is close to the raw time mean of the tide gauges—that is, averaging all the tide gauge time series together with no accounting for spatial weighting (cf., Church & White, 2011), but more importantly because it minimizes the differences in the regional trend analysis—even compared to CCAR—in section 3.2. There are some differences in the statistics presented if differing GMSL estimates are removed, but these are small and do not invalidate the results and conclusions presented. For the second comparison, in section 3.3, which calculates correlations and RMSD with the tide gauge and altimetry data, all data are locally detrended (by removing a linear least squares fitted trend line) and smoothed with a 5 year running mean so as to determine how well the data synthesis products match the low-frequency variability of the tide gauges' 5 year running means. This focuses on decadal climate variability by averaging out interannual variability. The analyses benefit from this averaging for the most part in the form of slightly improved correlations between products, gauges and altimetry, but the effect is small. Spatial smoothing was also applied (not shown), but had no substantial effect after the temporal smoothing, and is not included in the results.

Tide-gauge (TG) data were retrieved from the Permanent Service for Mean Sea Level (2015) that were marked as having no quality issues. Only those tide gauge time series were chosen that have sufficiently long records in the interval 1960–2007. Retrieved data were initially monthly mean data; existing data gaps of up to 24 months were filled in time by fitting climatological monthly averages into the gap, tilted by the "instantaneous" trend between the beginning and end points bordering the gap. Most gaps filled in this way are shorter than a year, with only a few gauges having gaps of up to 2 years. Gaps longer than 2 years were not filled. Monthly values were then averaged into annual means and shifted onto a $1^\circ \times 1^\circ$ grid (see below for the method); annual values on a $1^\circ \times 1^\circ$ grid are used in all analyses. Any gauge missing more than a total of 24 points, not necessarily consecutively, of annual data between 1960 and 2007 are dropped from further analysis. The three shortest remaining TG records have 31, 32, and 39 annual data points.

Gauges are corrected for the glacial isostatic adjustment effect using the same data set as used in the reconstructions (GIA; see Peltier, 2004), with the exception of CSIRO (Church et al., 2004), but no further local vertical uplift/subsidence estimates are corrected for. In addition, we applied an inverse barometer correction (Wunsch & Stammer,1997). However, noting that since we are only examining low-frequency variability and trends, we found this correction to have little effect on the results. Gauge locations are shifted up to one grid box during the comparison with reconstruction products to match a grid box where data is available for all (seven) analysis products being compared. Only in a few cases was this lateral displacement up to two grid boxes. After screening tide-gauge data and shifting their location to the nearest boxes with data from the seven products, there are 101 gauges available for the analyses presented here (see supporting information for further details).

Mean correlations between colocated tide-gauge data and reconstructions are calculated by first performing a Fisher transform on the individual correlations at each TG station (rendering them into normally distributed values), taking the mean, and then performing the reverse Fisher transform on the mean. Taking the median instead does not yield substantially different results. The ratio of RMS means of the products' and tide gauges' standard deviations are computed also on detrended values so that this is an estimate of the magnitude of the products variability compared to the tide gauges.

Satellite altimetry data—monthly means of delayed time, multimission sea surface height anomalies—were retrieved from AVISO (http://www.aviso.altimetry.fr/duacs/), at higher than $1^\circ \times 1^\circ$ resolution, and then were averaged into a $1^\circ \times 1^\circ$ grid of annual means.

Self-consistency comparisons were also conducted to explore how similar or dissimilar the products are between themselves without involving tide gauges or altimetry in the comparisons. These include maps of regional trends, low-frequency variability and basin-scale correlations of detrended data. Some extra

analyses involving EOFs are included in the supporting information, which highlight some of the shared variability in the Pacific and demonstrate that certain known features, such as the response to NAO variability, remain roughly robust even though, due to the details, quantitative metrics suggest strong disagreement in the Atlantic.

3. Results

3.1. Time Series at Tide Gauge Positions

For a first visual intercomparison, time series of all reconstructions are provided at the location of several tide gauges in Figures 2 and 3 and are displayed together with tide gauge time series. In these graphs, the

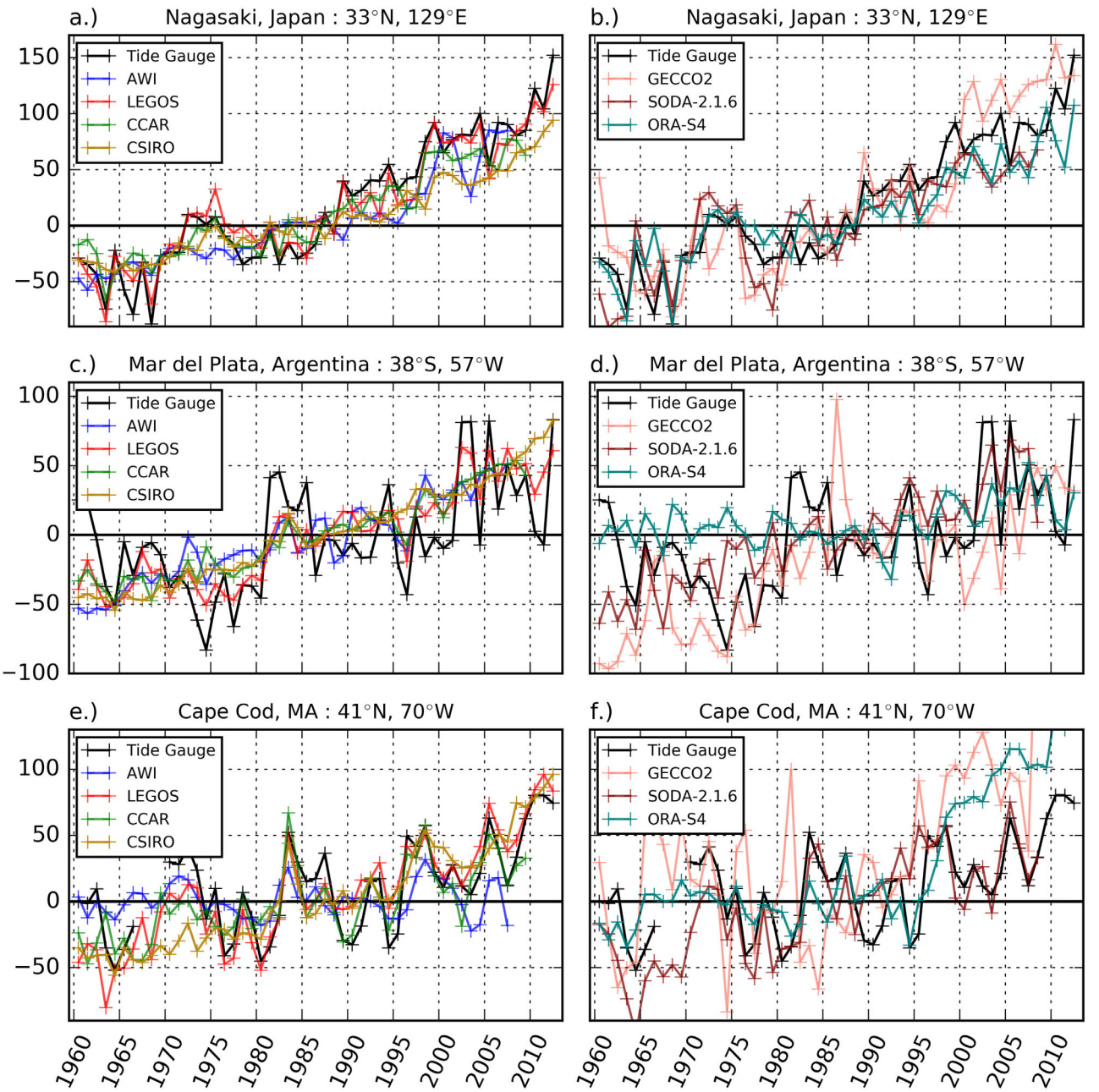

Figure 2. Time series of sea level relative to the 1960–1975 mean in mm. (left) Tide gauges (black) compared to the four tide gauge reconstructions. (right) Tide gauges compared to the three ocean data assimilation products at the same locations as those in the left figures. These data are unsmoothed and contain the global mean in order to directly compare the products visually.

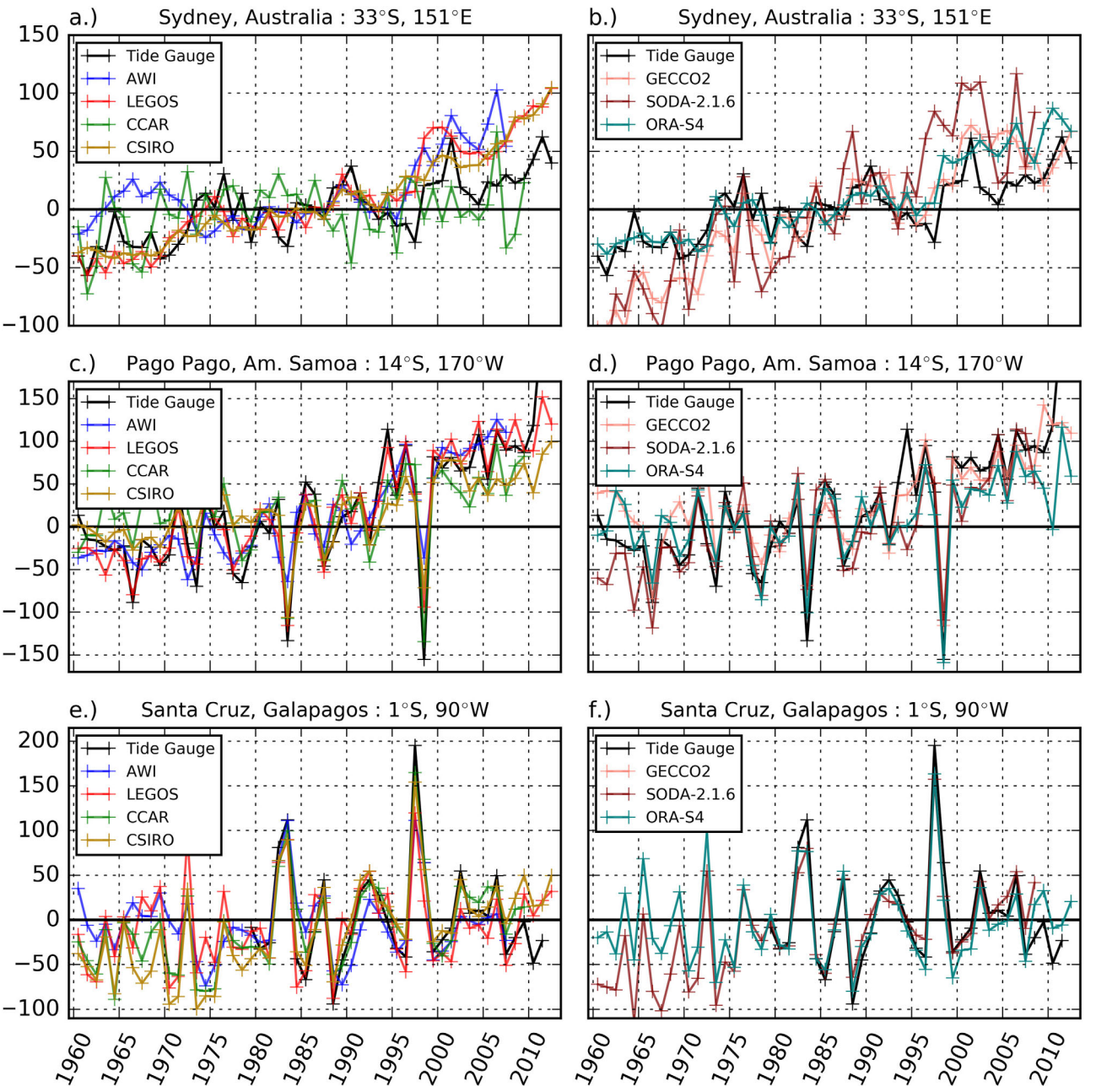

Figure 3. Time series of sea level relative to the 1960–1975 mean in mm. (left) Tide gauges (black) compared to the four tide gauge reconstructions. (right) Tide gauges compared to the three ocean data assimilation products at the same locations as those in the left figures. These data are unsmoothed and contain the global mean in order to directly compare the products visually.

global means are still present in order to directly compare the products visually in their rawest form. For the two ODAs which contain no estimated global-mean sea level which includes global thermosteric sea level, GECCO2 and SODA 2.1.6, the same mean removed from the tide gauges for trend comparisons in section 3.2 was added (see also section 2.4). For the most part, TGRs contain less variability than the tide-gauge data (left-side figures) or the ODA products, although in some cases they all together match the tide gauge variability very well (e.g., Figures 3c–3f). These particular locations were chosen because they include places where there is good agreement with the tide gauge and locations where there are substantial differences.

The TGRs often correlate fairly well with large signals in local SSH but do not reproduce the magnitude of the peaks, as in Figures 2e and 3c and 3e, but can completely miss the presence of smaller and shorter time

scale signals, as in Figures 2a and 2c. In contrast, the ODAs (right-hand figures) exhibit variability that is both smaller and larger than that of individual tide gauges, with the best matches to the tide gauges also occurring in the tropical Pacific (Figures 3d and 3f). The better results in the tropical Pacific, for all products, are probably due to the fact that the dynamics are well understood and captured for the most part in the data assimilation products and also are reasonably well sampled on interannual time scales by altimetry, and thus show up in the loading patterns for the TGRs. For the other gauges shown, the ODAs are not as consistent as the TGRs, and in some locations having considerably different variability and trends.

3.2. Regional Trends

Regional trends, minus any global-mean sea level (see section 2), are compared in maps of the 1960–1992 prealtimetry long-term trends (Figure 4), which reveal that the various products do not agree in magnitude or even sign of the estimated long-term trend over large areas of the ocean. In particular, two of the TGRs and two of the ODAs exhibit negative trends in the northern Indian Ocean, but the other products show either positive trends or mixed signals. Most products agree on a long-term midlatitude increase in SSH in the North Pacific Ocean, an increase that is the most robust signal between the products. Details differ in the exact spatial extent of this increase. There is some agreement of an SSH increase in the western midlatitude North Atlantic, as well as in the subtropical South Pacific.

Most other regions show a complete lack of consensus between all data sets, though some subsets of products show agreement between them. For instance, SODA 2.1.6 and ORA-S4 agree more in the Pacific and Indian Ocean. The PDO-like North Pacific pattern of AWI is supported by the strongly negative SSH trends off the western coast of Canada, although these TG trends may be too strong due to unresolved vertical land motion or a needed improvement in GIA estimates there (Slangen et al., 2014; Wöppelmann & Marcos, 2016). The pattern in the Indian Ocean matches the sparse tide gauges somewhat, but their coverage is insufficient to verify the reconstructions there. One curious feature is that AWI and CSIRO trends seem to have opposite patterns in the Pacific, and that these patterns are similar in structure to regional trend patterns in the altimetry data set. Having a similar pattern as the altimetry data could mean that they just imposed the pattern in the reconstructions or it could mean that this is real mode of oscillation. Since the two trend patterns have opposite sign, the former seems to be more likely, and it is also noteworthy that this pattern does not show up in the other five products.

Turning to the shorter-term trends over the satellite altimeter era (Figure 5), a different picture emerges with strong regional trend patterns from all products over most of the Pacific Ocean. This is a robust, large-scale feature which mostly agrees with the altimetric trends over the same time frame. The strong agreement in the Pacific is a theme which is seen in further figures and analyses and is likely due to the clear, large-scale patterns of SSH variability in the altimetry data there. Other coherent large-scale patterns in this trend analysis for the altimetric data (Figure 5h) appear in the low-latitude basins of the Atlantic and Indian Oceans, but it is also clear that the altimetric trends are more spatially noisy, i.e., with smaller scale features, in higher latitudes in all basins. There still remain differences in details, such as the extent of the negative trends in the extratropical Pacific. Other regions of disagreement include the Indian Ocean, and much of the Atlantic. Possible reasons for the disagreement between products in these regions include altimetry bias and the need for better vertical land motion estimates and are included in section 4.

The same regional product trends, plotted versus the colocated regional tide gauge trends, can be seen to lie close to zero for most products (Figures 6a and 6b). Because the global-mean time series has been removed from all products, this demonstrates that, over this longer-term period, the products' regional trends are mostly closer to their global trend than to the tide gauge trends. A cluster of AWI trends below the zero-line (x axis, Figure 6a) have lower trends than the tide gauges, but a few other locations also have stronger relative increases, which is a consequence of the strong spatial pattern of these 33 year trends seen in Figure 4a. ORA-S4 is another product that exhibits differences to the other products for this metric, as it has somewhat more positive trends than do the tide gauges for locations with weaker regional signals (Figure 6b). In general, TGRs have smaller median absolute differences to the tide gauge trends than do the ODAs (Table 2, column 1). Even so, over half of the TGR trends for each product differ from colocated tide gauge trends by more than 1 mm yr^{-1}, which is a substantial difference even for a regional trend of 5 mm yr^{-1} (Table 2, column 1).

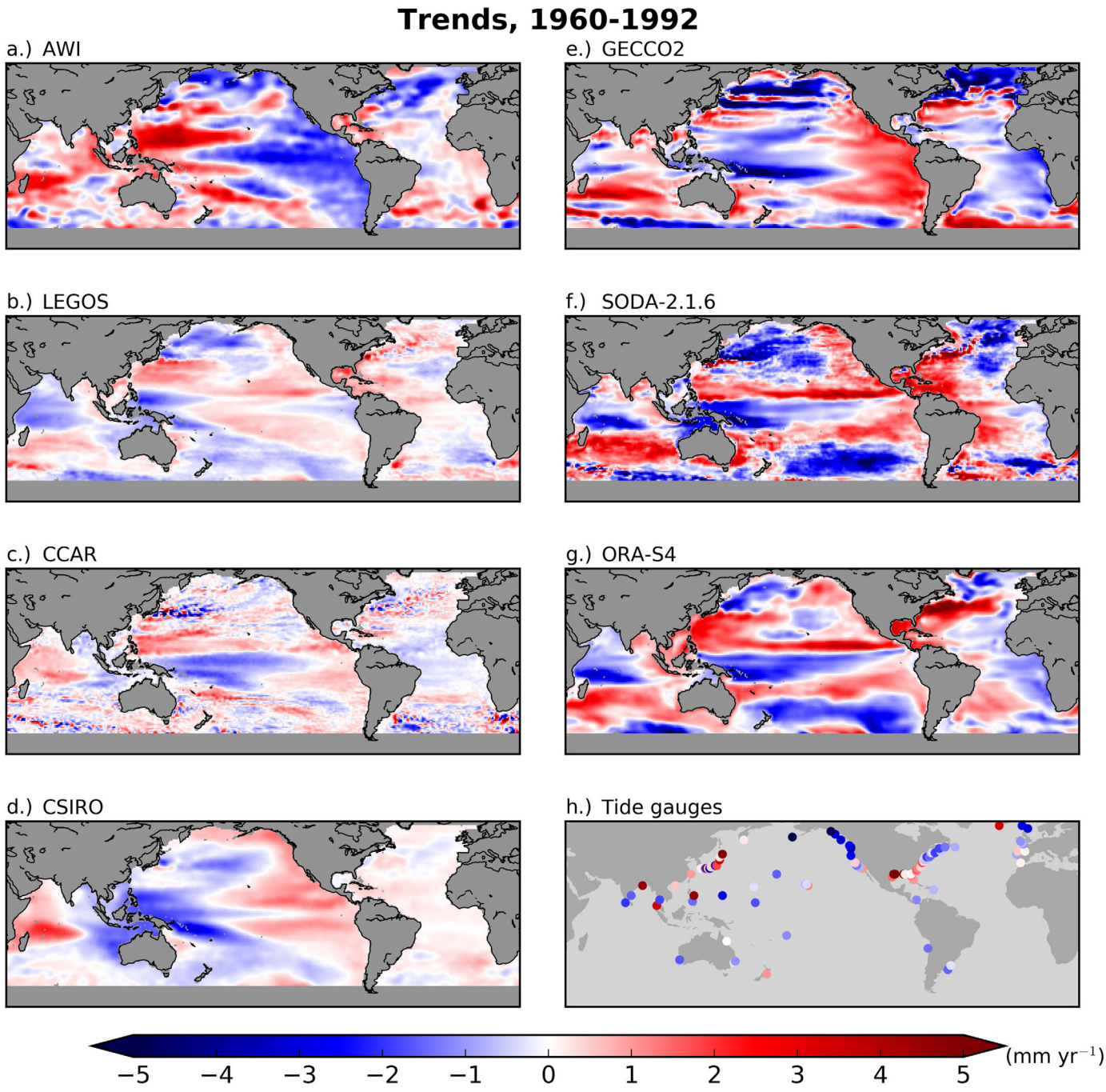

Figure 4. Trends calculated in each grid box between 1960 and 1992, in mm yr^{-1}. There is substantial disagreement between the products over much of the ocean, but a majority of products (in some cases, that is at most four of the seven) agree with the tide gauges off the coast of southern Japan, the Philippines, the Gulf of Mexico, and the northeastern coast of North America.

The altimetry-era trends present a more complicated picture, with many locations' regional trends differing significantly from the tide gauge trends (Figures 6c and 6d). For the TGRs (Figure 6c), there are certainly more locations with product trends within ± 1 mm yr^{-1} (within the grey band) of the tide gauge trend, even when the tide gauge trend is larger in magnitude, compared to the longer trends. There is still not much clustering around the grey line, while there is still some clustering around zero, indicating missing signals. But the differences are often greater in other locations, showing that the product trends in some regions are far off from the TG regional trend there. The TGR regional trends still do not deviate very far

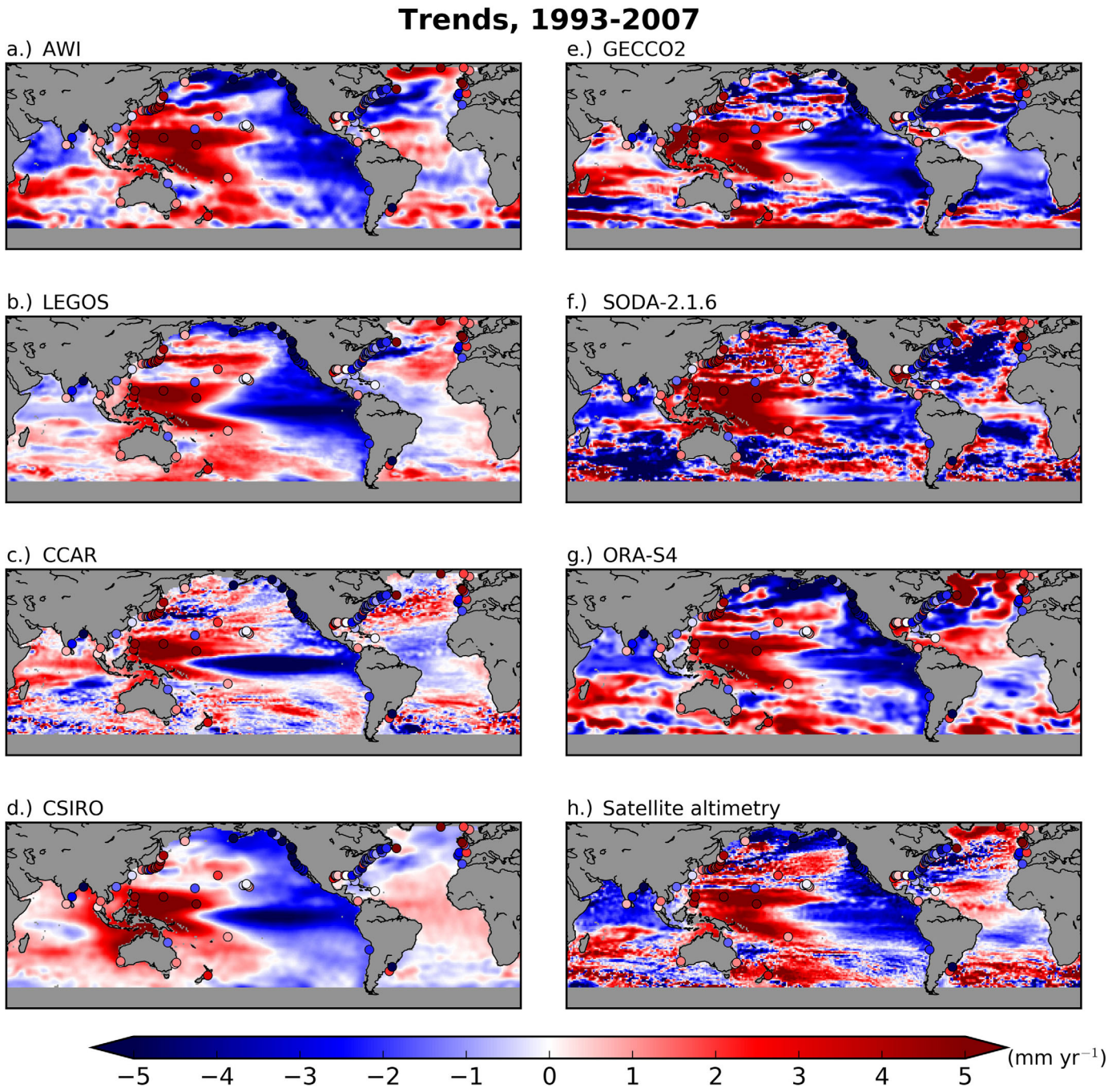

Figure 5. Trends calculated in each grid box between 1993 and 2007, in mm yr^{-1}. The agreement here is much better than in the previous figure for the longer time scale. For the most part, the TGRs agree well with the altimetry pattern in Figure 5h, although there are disagreements in the northern Indian Ocean, and the subpolar North Atlantic in the CCAR and CSIRO products.

from the global mean compared to the ODAs (Figure 6d). This is another expression of the weaker decadal variability in the TGRs relative to the ODAs, although this has the apparent benefit of the TGR trends being closer to tide gauge trends than those of the ODAs (Table 2, column 3). As for the 1960–1992 trends, some of the AWI regional trends are strongly negative compared with respect to tide gauges, though a few more also indicate stronger positive trends over this interval. For the ODAs (Figure 6d and Table 2, column 3), the situation for the altimetry-era trends is worse, as the spread is much larger than before, and there does not appear to be a tendency to only reproduce the short-term global-mean trend in the ODA products, i.e.,

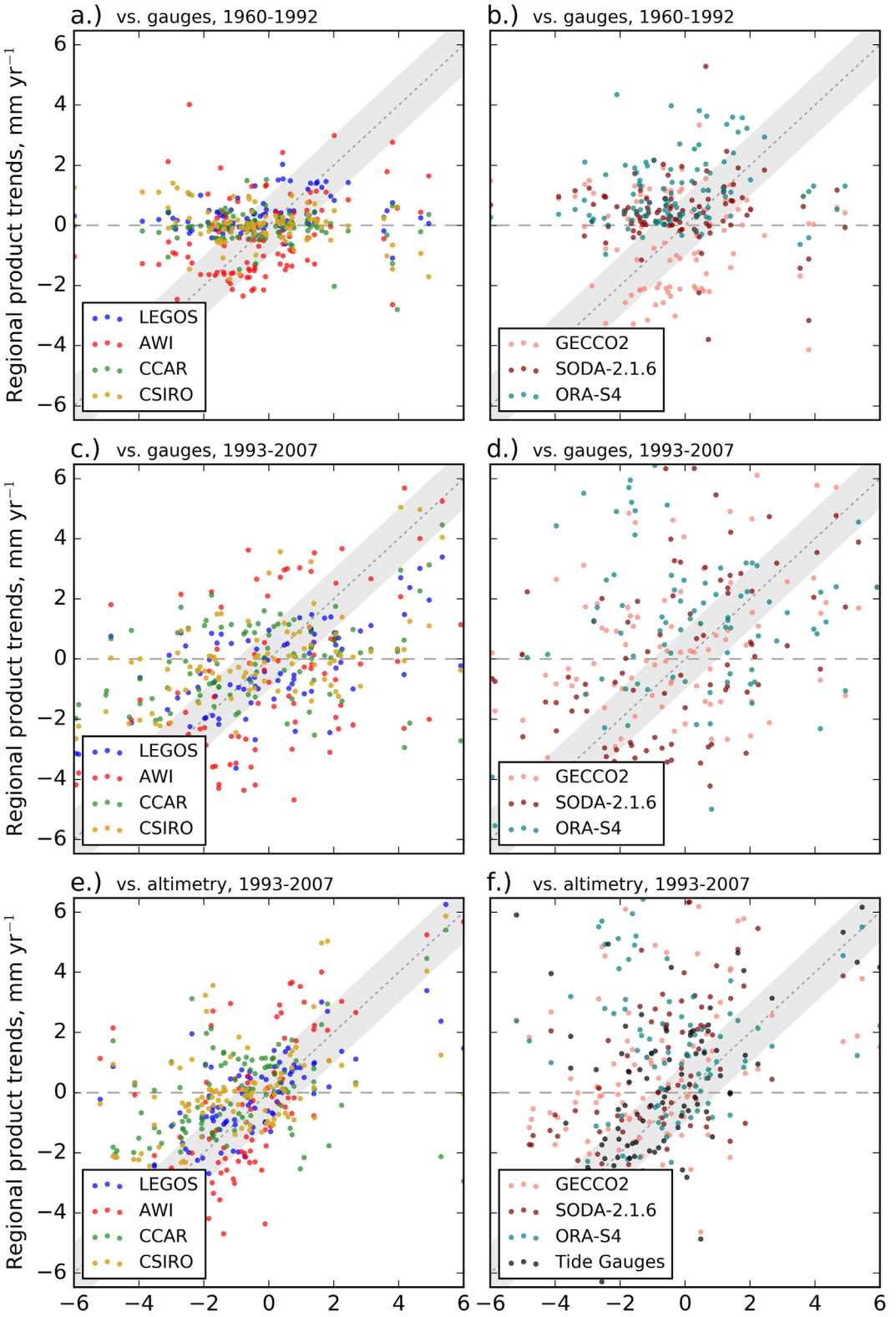

Figure 6. Regional trends, minus global-mean sea level Regional product trends (y axis) are plotted versus regional tide gauges trends (x axis) at the same location in a–d, and versus altimetry at the same (tide gauge) locations in e–f. Any regional product trend lying directly on the diagonal line $y = x$ represents perfect agreement with the tide gauge (or altimetry) regional trend. The ± 1 greyed out region around $y = x$ is a visual aid to show where product trends are within ± 1 mm yr^{-1} of the tide gauge (or altimetry) regional trend.

Table 2
Trend Difference Statistics (mm yr^{-1})

	Versus TG, 1960–1992		Versus TG, 1993–2007		Versus Sat, 1993–2007	
Product	MAV[a]	SDAV[b]	MAV	SDAV	MAV	SDAV
TGRs						
AWI	1.41	1.80	1.94	2.53	1.19	1.22
CSIRO	1.09	2.18	1.85	2.53	1.20	1.34
CCAR	1.29	2.05	1.75	2.59	1.31	1.53
LEGOS	1.10	1.85	1.44	2.35	0.81	1.03
TGR mean	1.06	1.94	1.40	2.47	0.53	1.18
ODAs						
GECCO2	1.73	2.13	2.60	2.64	2.42	2.27
ORA-S4	1.83	1.76	2.07	2.77	1.53	2.54
SODA 2.1.6	1.42	2.40	2.26	2.93	1.60	2.65
ODA mean	1.35	2.09	1.95	2.22	1.63	1.59
Ens mean					1.05	1.10
Tide gauges					1.43	2.66

Note. Trend differences are the product trend minus the tide gauge trend over the listed period, at each tide gauge location. Statistics are of absolute values to explore the magnitude of the disagreement regardless of direction of change (i.e., lower or higher sea level tendency).
[a]Median of absolute values of trend differences.
[b]Standard deviation of absolute values of trend differences.

tend toward zero regional trend, unlike for the TGRs. This is likely a consequence of greater decadal variability in the ODA products (see below).

From a comparison of all product regional trends directly with the satellite altimetry trends in the same locations as the tide gauges (Figures 6e and 6f), it is clear that products match the altimetric regional trends better than they do the tide gauge trends. In several locations, product trends differ from the altimetric regional trends by more than ± 1 mm yr^{-1}; nevertheless, all products match altimetric trends better than tide gauge trends (Table 2, cf., columns 3 and 5). We also note that for the TGRs, it is not surprising that regional trend anomalies, relative to the global-mean trend, would at least sometimes fall between the tide gauge trends and altimetric trends, since patterns of altimetric SSH variability are fitted to the tide-gauge data in some way. Therefore, some improvement might be expected. It is also likely that, for some stations, tide gauge SSH might simply be doing something different than the open ocean, due to local climate forcing or vertical land motion unaccounted for by the GIA correction.

For the ODAs, the improvement in matching altimetry instead of tide gauges could be an indication that these estimates represent ocean dynamics over this period, which is well captured by altimeter observations. In spite of this, the ODA results do not match the altimetric regional trends as well as the TGRs do. We also note that the tide gauge trends themselves (also displayed as black dots in Figure 6f) are not substantially better than the two best TGRs (LEGOS and CSIRO) and exhibit the lack of tendency toward the GMSL trend (i.e., zero regional trend), similar to the ODAs. It remains surprising to see the lack of agreement between GECCO2 and ORA-S4 with the altimetry better than this, given that they both assimilate altimetric sea level anomalies. However, GECCO2 does not assimilate satellite data close to the coast—within 200 km, and ORA-S4 assimilates with a larger uncertainty near the coast (Balmaseda et al., 2013). Also, sea level at coasts is not always the same as in the open ocean, even at low frequencies (e.g., Hong et al., 2000, also, see the following section). Finally, both products assimilate many additional data, and the assimilation procedure needs to satisfy all, and sometimes conflicting, constraints at once.

3.3. Low-Frequency Variability

The ability of the various products to reproduce the low-frequency variability found in tide gauge records and altimetric SSH is explored by applying a running mean with a sliding 5 year Hanning filter to the locally detrended data over the entire common period for the products, 1960–2007, and over the period from 1993 to 2014 for AVISO. All data in this section and the next were locally detrended in each grid box. The patterns of variability amongst the TGRs (Figures 7a–7d) are similar although with some clear differences,

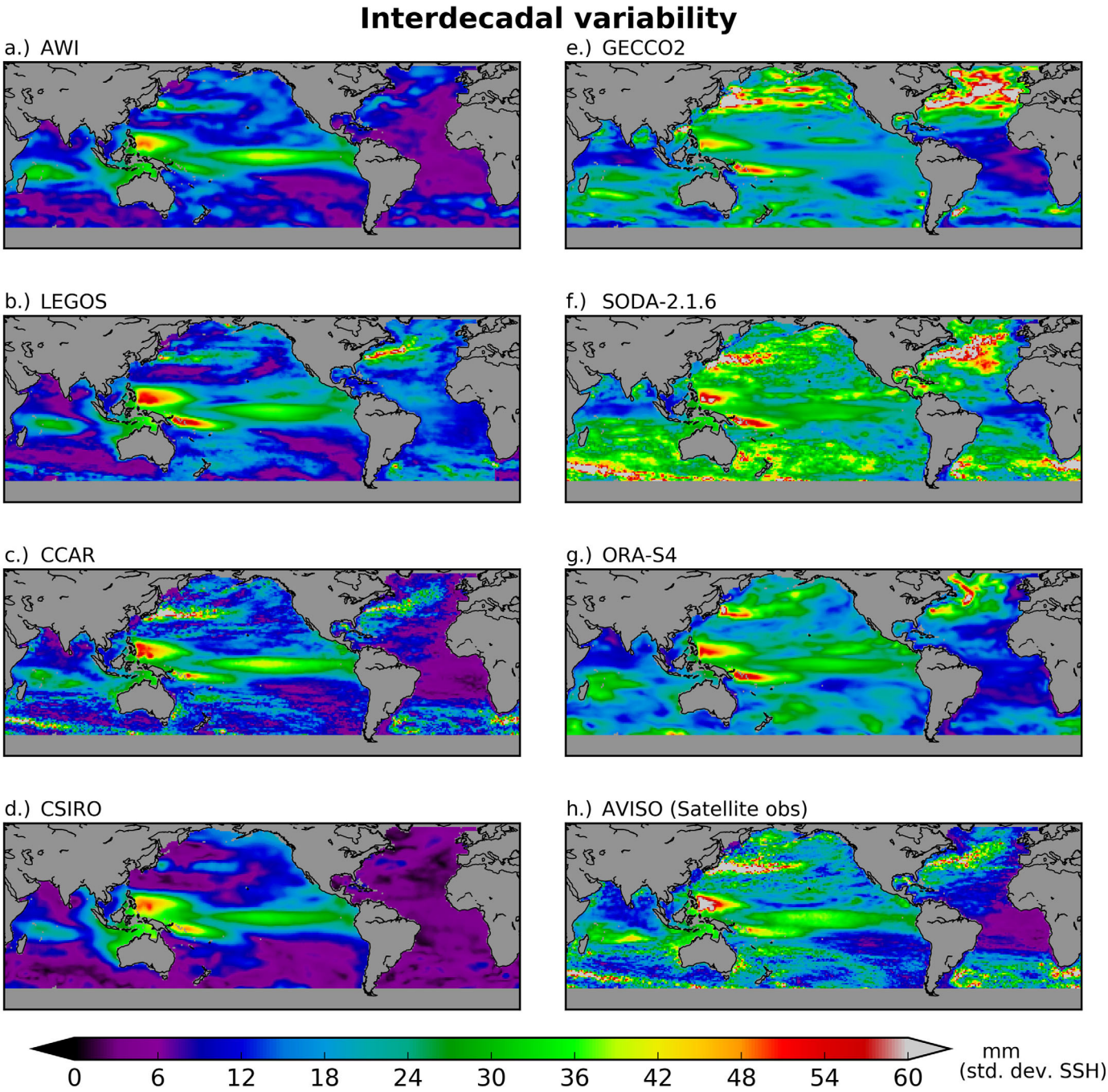

Figure 7. Interdecadal variability calculated from applying a 5 year running mean with a Hanning window to the annual data, and plotting the standard deviation in mm. The data coverage is over the entire common period for the products, 1960–2007, and over the period from 1993 to 2014 for AVISO. The tide gauge reconstructions all contain less variability at these frequencies than the AVISO observations or the data assimilation products, especially in the Atlantic oceans and the extratropics in general.

mostly agreeing to substantial decadal variability in the low-latitude Pacific and some minor variability in the North Pacific. Conversely, the variability in the North Atlantic varies between nearly 0 (CSIRO, Figure 7d) to over 30 mm (LEGOS, Figure 7b). The variability of the ODAs (Figures 7e–7g) is a little more similar to each other with stronger low-frequency variability than the TGRs, especially in the extratropics and the Indian and Atlantic Oceans, though with disagreement between them on the structure of the variability in these same regions. The ODAs also do not have the same consistency along the equatorial Pacific to the degree that the TGRs do. This tropical Pacific similarity among TGRs is certainly linked to the presence of substantial

decadal variability in the equatorial Pacific in altimetry data (Figure 7h), and the strong EOF modes which are generated from this data set (supporting information). The LEGOS data set shown here is an exception among the TGRs in that it is the mean of four different analyses, only one of which is directly based on altimetric patterns of variability. The presence of the reconstructions with input fields from ocean reanalyses in the LEGOS mean product is part of the reason for the higher variability in the Atlantic basins, another part being that the LEGOS reconstructions use EOF patterns of the input fields which were calculated, and subsequently fitted to tide gauges, per basin.

The comparison of low-frequency variability at the tide gauge locations is further explored by calculating the correlations of colocated time series which have been detrended and 5 year smoothed in all products, tide gauges, and altimetry. The results are then combined with the ratio of the standard deviations of these same time series and displayed as Taylor diagrams (Figure 8). Mean correlations in the North Pacific (triangles) are generally much higher than in the North Atlantic (plus signs), including with respect to AVISO (altimetry, black symbols in Figure 8c). This can also be seen spatially in Figure 9. For example, the tide gauge variability off the western coast of Europe is not represented very well by any of the products. The LEGOS reconstructions, however, reproduce the variability in both northern ocean basins much better than the other TGRs, particularly in the North Atlantic. The TGRs except for LEGOS, all of which are based on altimetry patterns, are not able to match the variability in the North Atlantic tide gauges as well as in the North Pacific. Two likely reasons for this seems to be that (1) North Atlantic patterns of variability during the altimetry period are not the same as in the prealtimetry period and (2) the global EOF patterns tend to fit the tide gauges in the North Pacific better than, or at the expense of, the North Atlantic since the variability in the North Pacific is more coherent over large scales and stronger (for further details, see the EOF analysis in the supporting information). The LEGOS reconstructions appear to benefit from the two main differences with respect to the other reconstructions: that three of the LEGOS reconstructions use EOF patterns based on the longer time scale data from data assimilation products and that LEGOS fits the Atlantic Ocean separately from the Indo-Pacific ocean. This latter difference leads to a larger number of degrees of freedom for the fit, and may lead to overfitting (see below). The better correlation in the North Pacific is also noteworthy since there is no analogous improvement of regional trend differences in the North Pacific versus the North Atlantic for any of the products (not shown). In addition to these central tendency results, some specific tide gauges do not match altimetry due to the short spatial structure of some coastal dynamics (Vinogradov & Ponte, 2011).

Total mean correlations (squares) are generally higher over the satellite altimetry era (Figures 8c and 8d) than over the 33 year prealtimetry period (Figures 8a and 8b), though the increase is larger for the TGRs than for the ODAs. As with the trends above, this is not entirely surprising, given that the open ocean patterns of SSH variability during the satellite era is well measured, but only the tide gauges themselves are available for describing SSH variability before this period. The ODA correlations are not much improved over the more recent altimetry period. The lack of improvement in ODA correlations are likely partially linked to coastal processes that affect gauges, but which are poorly captured in some reanalyses (e.g., Chepurin et al., 2014; Piecuch et al., 2016; Sturges & Douglas, 2011), and partially to locations which are not simply driven by the large-scale climate signals that reanalyses do capture well, e.g., in the Pacific where the ENSO signal is not as strong (Chepurin et al., 2014). The TGRs have less variability than the TGs over both periods, though they match the TG data slightly better over the altimetry period (Figures 8c and 8d). Two of the three ODAs, GECCO2 and SODA 2.1.6, have nearly the same amount of variability as the tide gauges in both periods. This is another example which shows that the TGRs underestimate variability, as was seen in Figure 7. The comparison of altimetry time series with colocated tide gauges are shown as the black squares in Figures 8c and 8d (same values in both figures). It is unclear exactly why the AWI and CSIRO TGRs fail to capture North Atlantic tide gauges as well as altimetry does (Figure 8c).

When comparing the products directly to the altimetry data, all resulting correlations are better than those obtained from the comparisons to tide-gauge data (Figure 8e and 8f). Still, in all products, some locations continue to be very poorly correlated with both the local tide gauges and altimetry. The squares are the mean correlation and RMS variability ratio only for where the tide gauges are located, as before for comparison. The mean correlations and RMS mean variability ratios for the North Pacific and Atlantic (triangles and plusses) are still the values computed only at the tide gauge locations in those basins. Like the trend differences described above, the TGRs match the altimetry better presumably because the simultaneous fit to

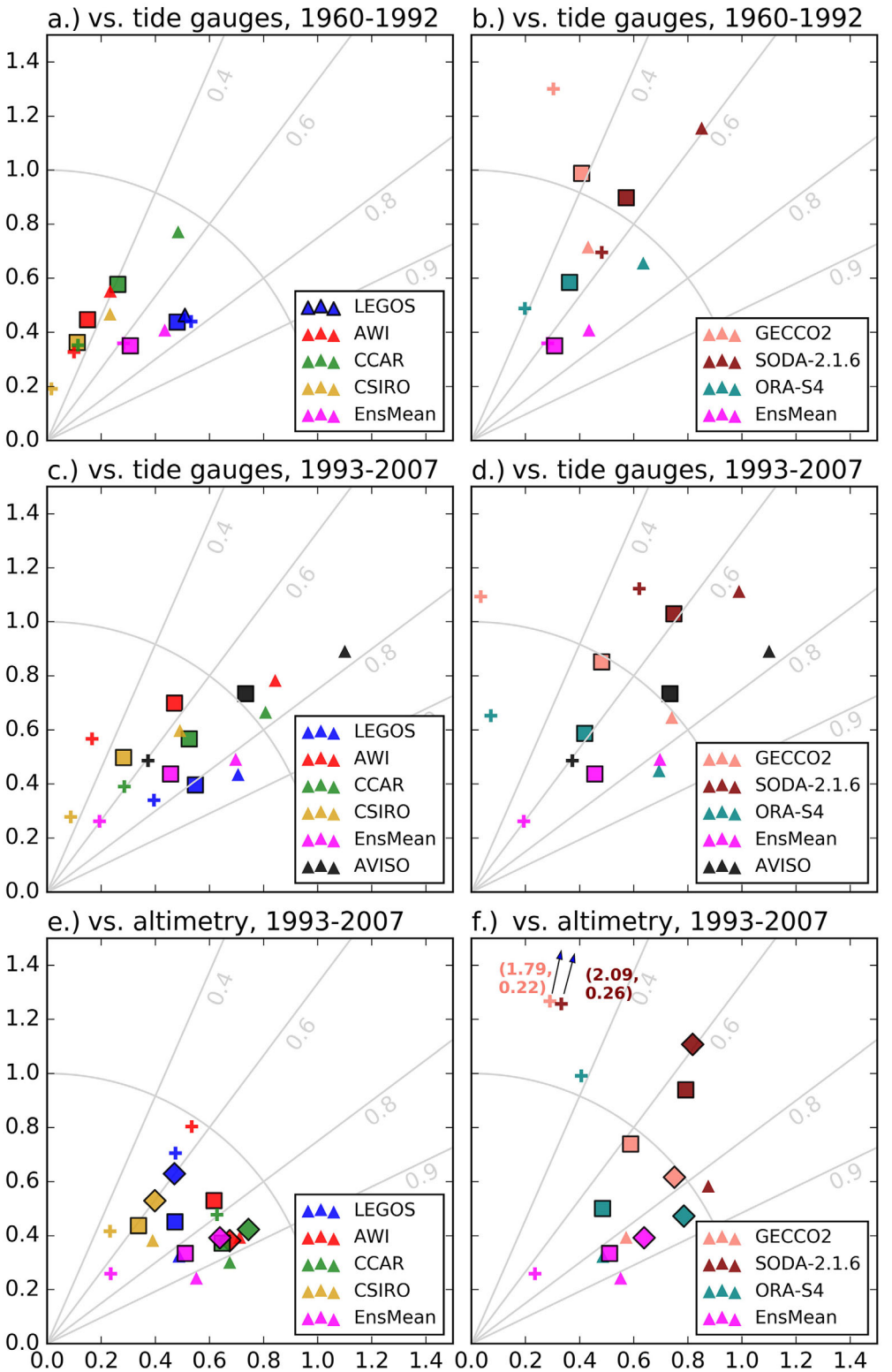

Figure 8. Taylor diagrams of correlations (the angle) versus the ratio of standard deviations (the radius) between the various products and the tide gauges and altimetry, all from detrended, 5 year-smoothed data. A radius greater than 1 means that the product's standard deviation is larger than that of the tide gauge or altimetry at that location. Squares represent mean correlation and RMS mean variability ratio for the products over all TG locations, even for the comparison to altimetry in e and f. Triangles and plus signs are the means over only the North Pacific and North Atlantic, respectively. The diamonds in Figures 8e and 8f are the means over all locations for a global comparison of the products to altimetry (i.e., not just at TG locations).

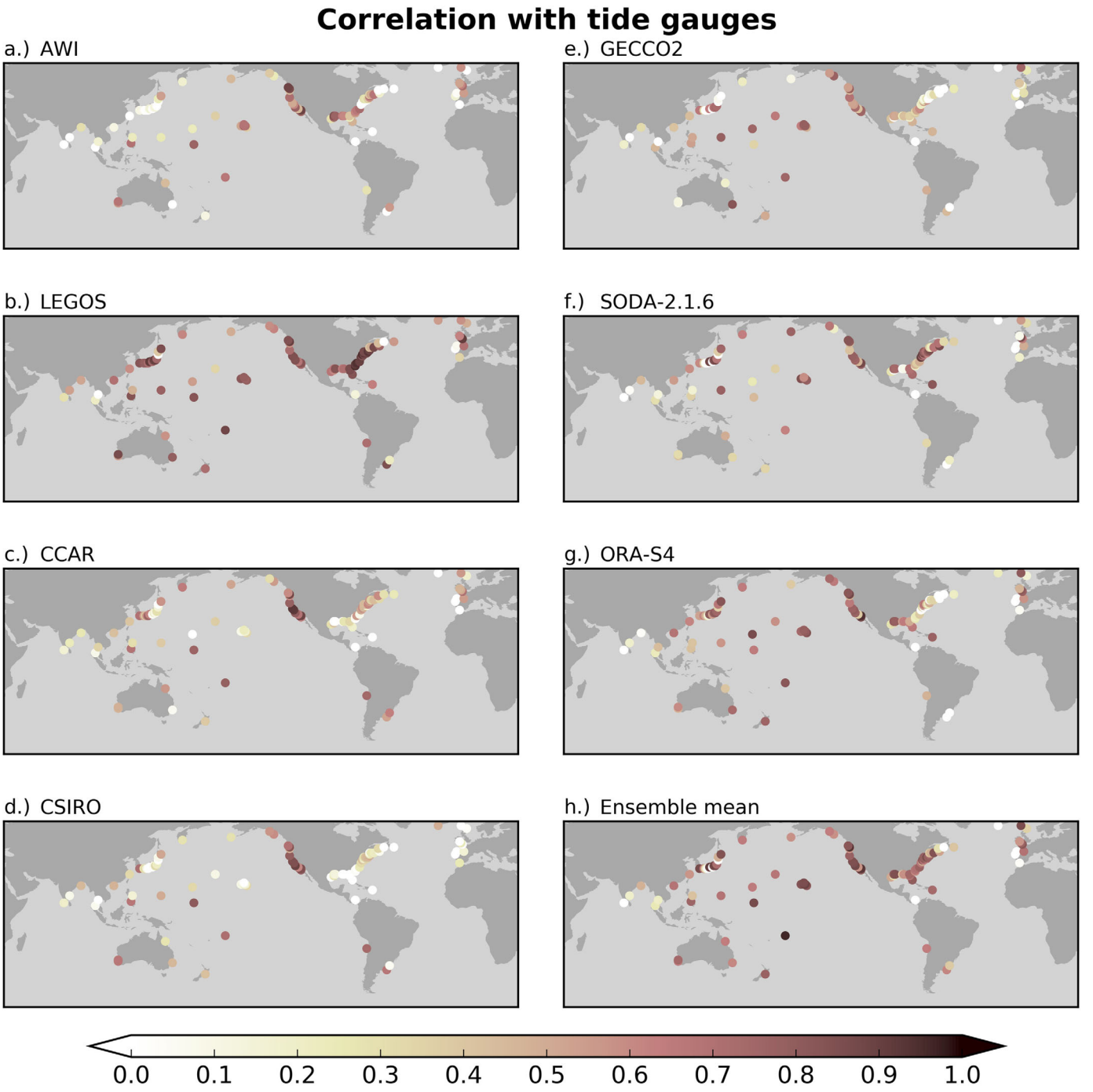

Figure 9. Correlation with tide gauges, prealtimetry. Tide gauge locations showing the correlation with the listed product over the period 1960–1992. The correlation is calculated after a 5 year running mean has been applied to each time series, and all have been detrended.

multiple tide gauges will not reproduce the full variability at any of the tide gauges, whereas the fitted variability patterns stem in some manner from altimetry or (in the case of some of the LEGOS reconstructions) large-scale patterns from ocean reanalyses. The LEGOS mean reconstruction, which performs better than the other TGRs in nearly all the previous metrics, does not compare so well to altimetric low-frequency variability at the tide gauge locations and compares even less well in global-mean correlation (blue diamond versus blue square). This is possibly due to the use, in LEGOS reconstructions, of a fixed number of tide gauges through the entire period, which means that additional tide gauges that are available during the

altimetry era, and are used by the other reconstructions, are not used to improve the LEGOS products' performance over this later period. Conversely, the AWI product, which did not perform as well as most of the other TGRs in previous metrics, performs very well in reproducing the low-frequency variability of the altimetric data over the whole ocean (red diamond).

The reanalyses themselves (ODAs) also match altimetric variability better than tide gauge variability at these same locations (Figure 8f versus Figure 8d). And, when taken over all ocean grid boxes where the products can be compared to altimetry (i.e., not just at the tide gauge locations), the "total" mean correlation (the diamonds in Figures 8e and 8f) can differ substantially from the mean correlation estimated only from tide gauge locations (which can also be seen spatially in Figure 10). GECCO2 and ORA-S4, which do assimilate the altimeter data, also compare much better to the altimetric data than they did to TG data, even more when considered globally (diamonds in Figure 8f). The considerable improvement in the two ODAs which assimilate altimetric sea level anomalies, GECCO2 and ORA-S4, over the whole ocean (diamonds) versus only at tide gauge locations (squares) highlight how much better these products match altimetry in the open ocean versus at the coastlines, as also suggested earlier in section 3.2. Looking at the LEGOS correlation with altimetry (Figure 10b), the fact that these correlations are generally weaker than those with respect to the tide gauges (Figure 9b) suggests that the LEGOS method of fitting the tide gauges to EOFs estimated per basin may lead to overfitting at tide gauge locations at the expense of capturing altimetry variability.

Correlation with the PDO climate index, with all data and the index detrended, shows high correlation over most of the Pacific Ocean, with some good agreement in the Indian, and mild agreement in the Atlantic (Figure 11). This is another view into the decadal variability that exists in the products, with this being the only real coherent pattern of low-frequency variability between the products. This level of correlation is not seen for the NAO (not shown, cf., North Atlantic EOFs in the supporting information).

3.4. Spatial Robustness of Results

One further way to quantify the robustness of these results is to compare the extent to which these products reproduce the same results between themselves, and not just how they compare to tide gauges. This test is performed on the detrended time correlation, per grid cell, over all ocean regions in the two basins best constrained by existing tide gauges, i.e., the North Pacific and the North Atlantic. These are the two basins which have the best likelihood of having reproducible results between products, as they are the best sampled regions in the prealtimetry era. The results of these tests can show where results are more sensitive to the methods used, and where they are less so. This decouples the discussion from comparisons to tide gauges and altimetry and examines the products' quality independently from instrumentation.

Mean correlations between products in the North Pacific are much higher than in the North Atlantic (Table 3), which is somewhat to be expected given the results previously stated and as shown in Figures 8, 9, and 11. Most of the correlations listed in Table 3 for the North Pacific are fairly close together, with only marginally better correlations between LEGOS and the ODAs compared to the other products versus ODAs. Also interesting to note, are the mostly higher correlations among the ODAs (Table 3, boldfaced values) compared to the correlations among the TGRs (Table 3, italicized values).

This difference becomes stark when looking at the results for the North Atlantic. Most mean correlations among TGRs are below 0.20, which is another indication of how poorly similar methods can reproduce similar patterns. The reason for this is not so clear, but is probably due to one or both of the following factors: tide gauge variability does not represent open ocean variability in the North Atlantic and thus cannot be fit to altimetry-based open ocean SSH patterns; or, altimetry has not been running long enough to capture lower frequency modes of variability in this region, which is needed for fitting tide gauges to produce the correct open ocean variability. Yet another possibility is that sea level in the North Atlantic does not exhibit large dominant modes of variability like in the Pacific and is poorly captured by a truncated set of EOFs.

4. Discussion

While several sea level reconstructions exist for the past 50–60 years, all products show significant deviations from each other in their regional sea level variability and trends. Only in some regions do the products appear to agree better over the entire period. However, even during the well-observed satellite altimetry

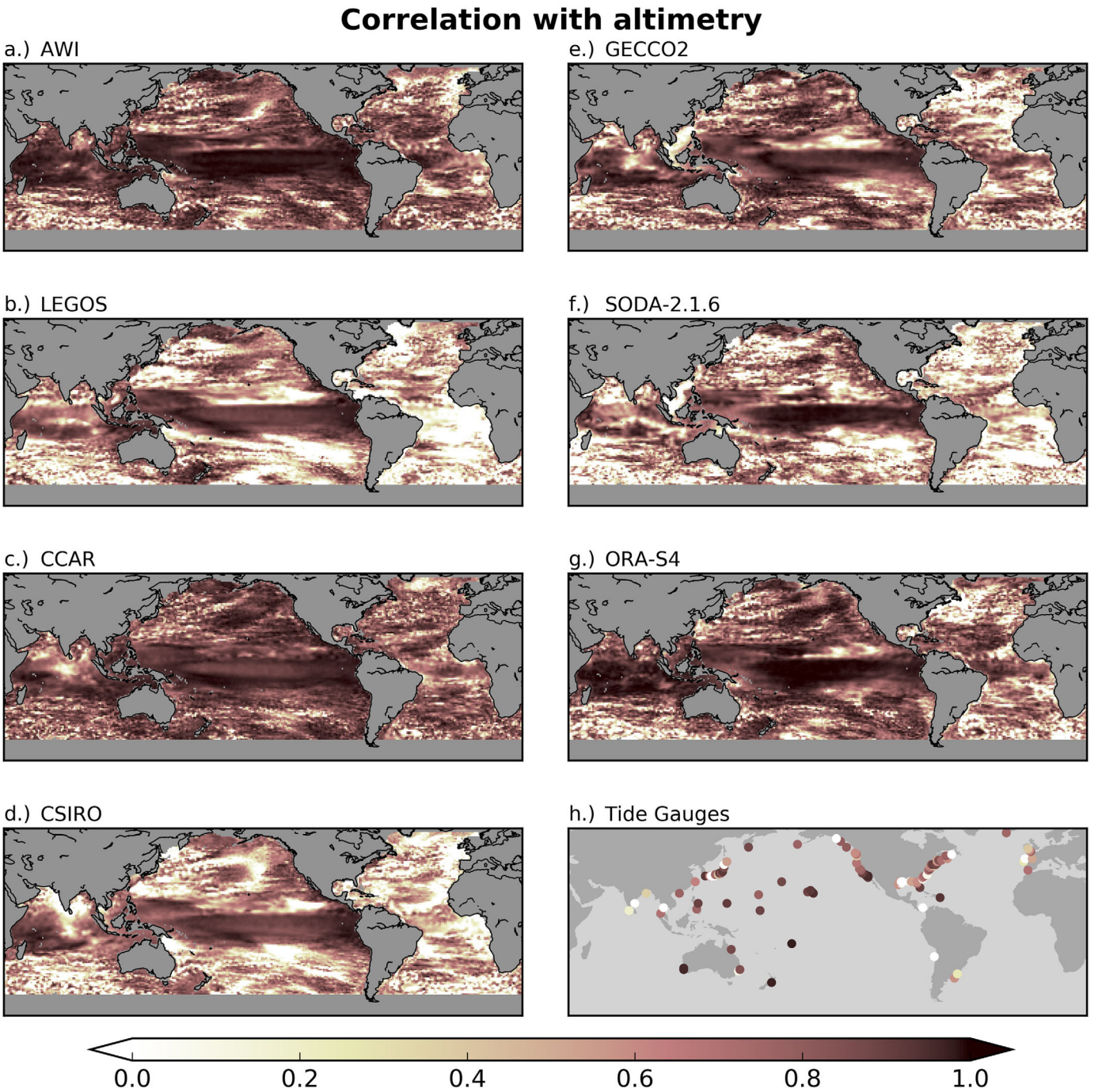

Figure 10. Correlation with altimetry. The correlation of satellite altimetry data with the listed product over the period 1993–2007. The correlation is calculated after a 5 year running mean has been applied to each time series, and all have been detrended.

era, they only agree particularly well with each other in the Pacific Ocean, and parts of the Indian Ocean (which covary with the Pacific). And this agreement is largely restricted there to ENSO and PDO patterns of interannual and decadal variability. In contrast, the variability in the products is not well correlated with tide gauges, or each other, outside the Pacific.

With respect to regional trends, i.e., the anomaly relative to the global-mean, trends inferred from tide-gauge data are generally larger in magnitude (more positive or more negative) than the product trends found in reconstructions or even altimetry, suggesting that remaining uncertainties exist also in the

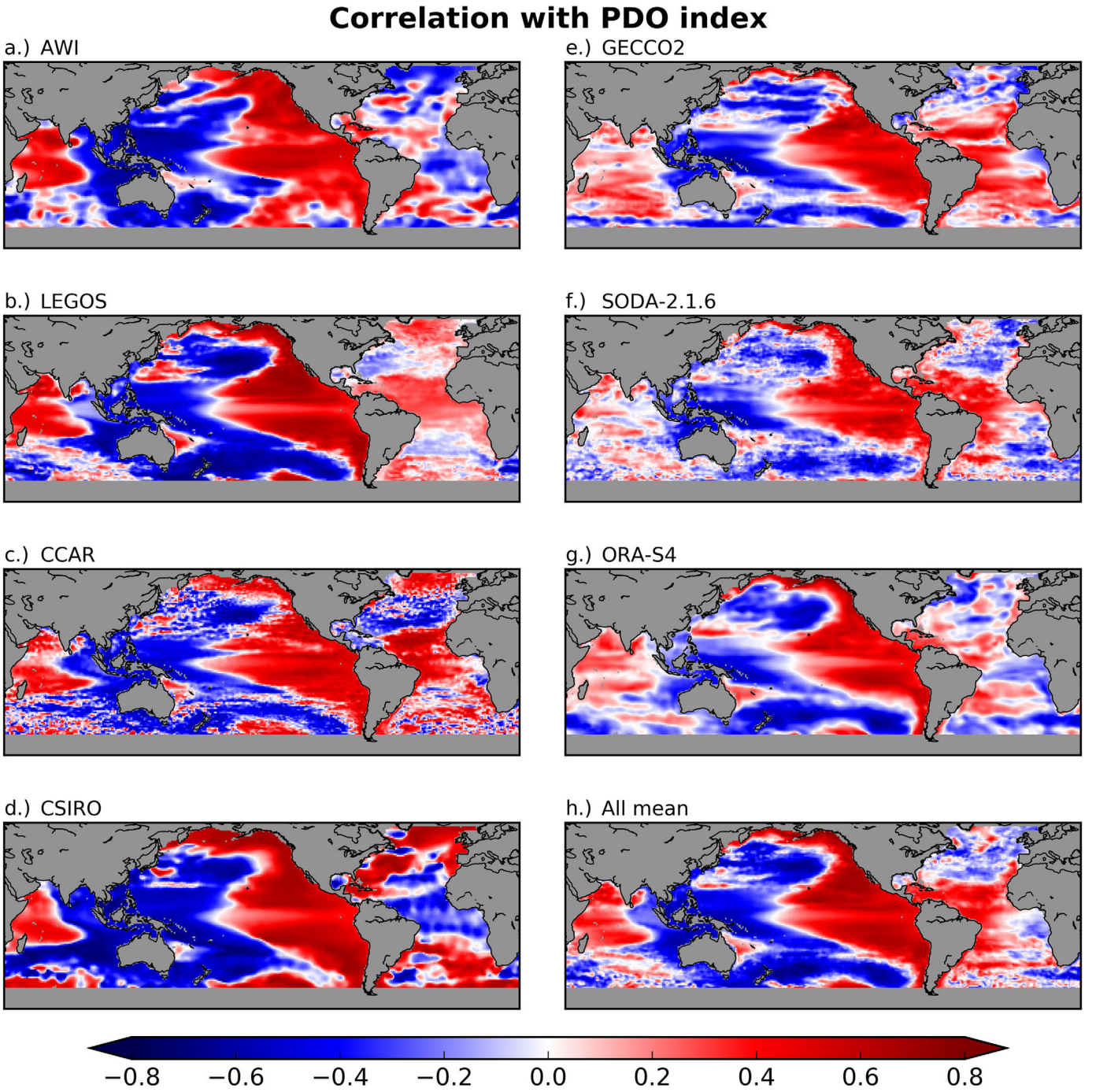

Figure 11. Correlation of the products with the PDO index. There is good agreement between all products for this climate mode, but differences in details are present and substantial in some regions. This is the only climate mode index for which all the products agree this well, or for which all have large-scale, good correlations.

observational data record. The reason why tide gauge trends can deviate from open ocean sea surface trends measured by altimetry fundamentally depends on the way local vertical land motion at the tide gauge location is estimated and can be corrected in the tide-gauge data, along with any other residual trend in corrections (King et al., 2012). Improvements in our ability to infer and correct local vertical land motion are an area of ongoing research (Santamaría-Gómez et al., 2014; Wöppelmann & Marcos, 2016), and might require that all tide gauge stations come equipped with appropriate GPS receivers, which may be able to better constrain local estimates of GIA and other sources of VLM. Correcting the gauges for these

Table 3
Correlations Between Products[a]

	AWI	CSIRO	CCAR	LEGOS	GECCO2	ORA-S4
North Pacific (0°N–65°N)						
TGRs						
CSIRO	*0.62*					
CCAR	*0.63*	*0.73*				
LEGOS	*0.58*	*0.73*	*0.69*			
ODAs						
GECCO2	0.52	0.56	0.61	0.66		
ORA-S4	0.54	0.61	0.64	0.68	**0.74**	
SODA 2.1.6	0.51	0.57	0.58	0.64	**0.69**	**0.80**
North Atlantic (0°N–65°N)						
TGRs						
CSIRO	*0.11*					
CCAR	*0.31*	*0.10*				
LEGOS	*0.12*	*0.11*	*0.14*			
ODAs						
GECCO2	0.27	0.02	0.10	0.26		
ORA-S4	0.33	0.00	0.20	0.39	**0.47**	
SODA 2.1.6	0.28	0.06	0.20	0.37	**0.40**	**0.53**

[a]Calculated from detrended time series of 5 year running mean annual sea level anomalies. The correlations listed are mean correlations over all grid boxes in the named basin, i.e., over all ocean regions in the basin, calculated as before and described in section 2.4. The italicized values are the correlations of TGRs with TGRs and the boldfaced are of ODAs with ODAs, to highlight these results for easy inspection.

motions before fitting them to RSOI patterns from altimetry may help improve regional trend estimates perhaps to a large degree, as errors have been found as large as ± 2 mm yr^{-1} in long-term linear trends. Remaining altimetric biases in regional trends can also exist from limitations in tide gauge coverage and the need for better vertical land motion estimates (Watson et al., 2015); these biases should be corrected with the latest estimates to ensure that the resulting reconstruction contains as little overall bias as possible. A continuation of the ESA CCI project (Ablain et al., 2014) could therefore pay specific attention to the estimates of regional sea level trends, especially close to coast lines.

But it is important to recognize that regional trends can also be incorrectly reconstructed due to insufficient resolution of lower frequency modes of variability in the shorter altimetry record. We find that the North Atlantic in particular is poorly reconstructed in general when comparing the correlations between products and to the tide gauges, and especially in the prealtimetry period. The LEGOS TGR mean product performs better in the North Atlantic than the other products in the comparisons to tide gauges, due in part to the per-basin EOF patterns used in these specific reconstructions (all four LEGOS reconstructions perform well there). A comparison between the LEGOS reconstructions over the prealtimetry era (not shown) reveals that two of the three which are based on the longer time scale patterns from ODAs (the ORA-S4 and SODA based reconstructions) have slightly higher median correlations to tide gauges in both the North Pacific (0.71 and 0.68 versus 0.65) and North Atlantic (0.75 and 0.7 versus 0.65) than for the altimetry-based reconstruction. Since the ODAs already compare less well to the tide gauges than does altimetry over the altimetry era (Figure 8d), this suggests that there is some benefit to the longer time frame of the ODAs' sea level data, which could generate sea level variability not accounted for in the shorter altimetry record.

One recommendation for future efforts emerging from this study would be to test the performance of TGRs constructed separately for each basin. Then methodological differences, such as cyclostationary EOFs or the approach of fitting gauge PC time series to altimetry patterns, would provide results which could be tested for robustness of this per-basin approach. The goal could be to specifically examine open ocean SSH variability in the North Atlantic to see if there is some improvement in correlations between treatments in the ocean away from the tide gauges. A second possible methodological change could be to remove an estimate of GMSL initially from all input data in an effort to specifically reconstruct regional trends, alongside additional corrections of local vertical land motion as suggested above. However, an important caveat remains, in that we would still not be certain that even robust features have captured historical open ocean

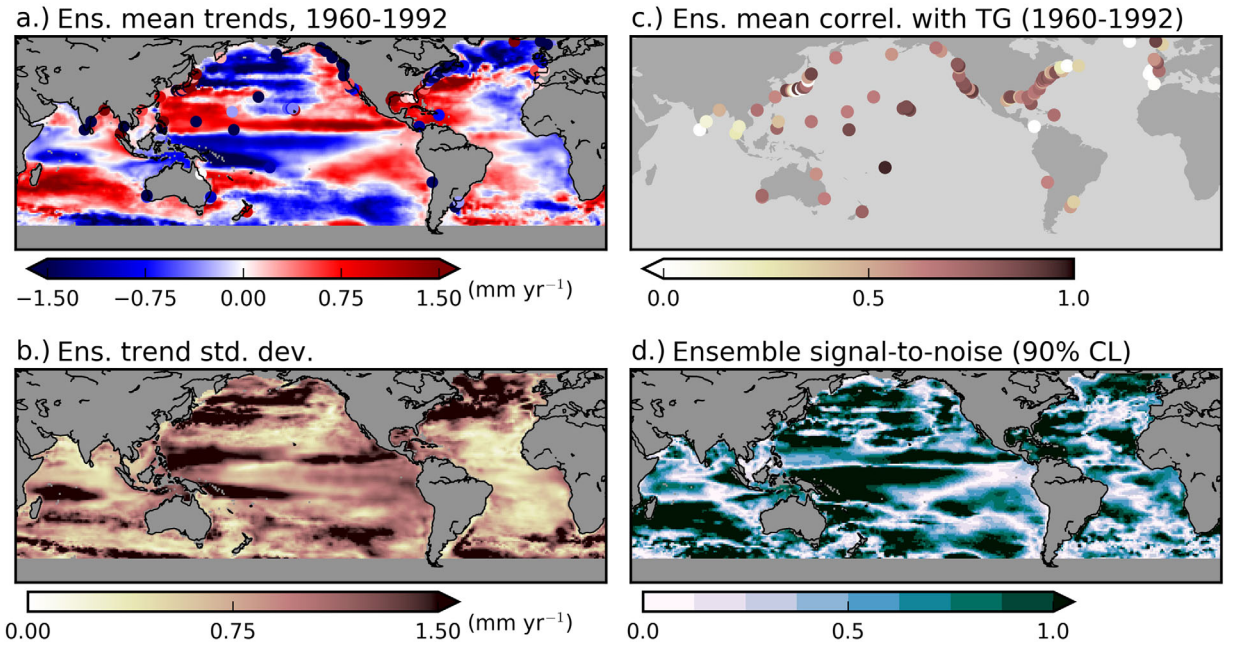

Figure 12. Ensemble statistics. (a) Regional trends of the ensemble mean from 1960 to 1992 in mm yr^{-1}, relative to the global-mean trend. (b) Ensemble spread (standard deviation) of the seven products' 1960–1992 trends, in mm yr^{-1}. (c) Correlation of ensemble mean detrended time series with detrended tide-gauge data, after a 5 year running mean was computed. (d) Signal-to-noise ratio (Figure 12a divided by Figure 12b scaled to the 90% CL); greater than 1 is 90% significant.

SSH variability patterns particularly well. There also continues to be the possibilities that either (1) altimetry has not been around long enough to capture lower frequency North Atlantic variability modes of importance or (2) the tide gauges are simply doing something different than the open ocean in that basin. In the latter case, various data processing approaches may be needed since gauges can have different reasons for mismatching open ocean variability, including riverine variability, vertical land motion, and local wind forcing.

An important question remains whether uncertainties in regional SSH trend estimates contain random errors, and whether an ensemble mean of the existing reconstructions could provide a better estimate. A similar approach is performed in the context of ocean reanalysis (Balmaseda et al., 2015; Storto et al., 2015), and with the LEGOS products (Meyssignac et al., 2012a). Testing it here in the context of SSH reconstructions, it appears that averaging the products will improve the performance of the result with respect to the tide gauges, as was seen for the regional trends in Table 2 (MAV columns for TGR mean, ODA mean, and Ensemble mean). The ensemble mean trend (Figure 12a) still shows some resemblance to the shorter altimetry trends in the Pacific, which results partly from the strong AWI trends in this basin over this long period. In general, the ensemble mean product does not show a conclusive improvement relative to all tide-gauge data (circles in the figure), not unlike for the individual products. We also note that the ensemble spread of the trends exhibits greater uncertainty in the western low-latitude Pacific and in western boundary current separation regions in the North Pacific and North Atlantic (Figure 12b). The subpolar Atlantic and eastern tropical Pacific also have significant spread of nearly 1 mm yr^{-1}, which is half the size of the global-mean trend, and a substantial part of the regional tide gauge trends.

Does this ensemble mean product provides a better product in the open ocean than the individual products? It appears that the decadal-scale correlations to detrended tide gauges shows improved low-frequency variability relative to most of the products (Figure 12c). However, an independent test of the accuracy of the ensemble mean over the open ocean is still needed. When we compared the ensemble mean of all seven products to the ensemble mean of just the TGRs and that of just the ODAs, we find that the TGR ensemble mean has nearly the same mean correlation with the prealtimetry tide-gauge data as does the ensemble mean (0.65 compared to the full ensemble mean's value of 0.66), whereas the ODA ensemble mean performs somewhat less well (0.55). This also shows that the TGRs still match the tide

gauges better than the ODAs, although it also still does not describe how well these products reproduce the real variability in the open ocean.

At present time, a significant limiting factor in making further progress is the lack of long altimetric time series, preventing us from estimating accurate long-time-scale SSH variability patterns. Only the future can remedy this problem, underpinning the need to continue indefinitely high-quality altimetric measurements. Tide gauges need to also be maintained into the future, and continuing efforts are needed to rescue historical tide-gauge data.

From our study, we conclude that presently a coherent description of regional sea level changes prior to the altimeter era only exists to a limited degree. The reconstructions of past sea level presented here show substantial uncertainties and do not allow robust conclusions about prealtimetry regional sea level variability and change patterns over much of the ocean; there is some agreement in the detrended North and tropical Pacific data related to climate modes. Finding a robust way to infer improved estimates of past regional sea level changes and variability remains a subject of fundamental research.

Acknowledgments
This work was supported in part through the DFG funded CliSAP Excellence Cluster of the University of Hamburg, through the BMBF (Federal Ministry of Education and Science) Project RACE, and through a DFG funded special research effort on regional sea level (SPP1889). The altimeter products were produced by Ssalto/Duacs and distributed by Aviso, with support from Cnes (http://www.aviso.altimetry.fr/duacs/). The authors acknowledge the ERA4CS project ECLISEA for supporting the development and production of the LEGOS 2D sea level reconstructions. The data sets explored in this paper are available from repositories listed with the references cited in Table 1 in the References section.

References

Ablain, M., Cazenave, A., Larnicol, G., Balmaseda, M., Cipollini, P., Faugère, Y., . . . Beneviste, J. B. (2014). Improved sea level record over the satellite altimetry era (1993–2010) from the climate change initiative project. *Ocean Science Discussions*, *11*, 2029–2071. https://doi.org/10.5194/osd-11-2029-2014

Ballu, V., Bouin, M., Siméoni, P., Crawford, W. C., Calmant, S., Boré, J., . . . Pelletier, B. (2011). Comparing the role of absolute sea-level rise and vertical tectonic motions in coastal flooding, Torres Islands (Vanuatu). *Proceedings of the National Academy of Sciences of the United States of America*, *108*(32), 13019–13022. https://doi.org/10.1073/pnas.1102842108

Balmaseda, M. A., Hernandez, F., Storto, A., Palmer, M., Alves, O., Shi, L., . . . Gaillard, F. (2015). The Ocean Reanalyses Intercomparison Project (ORA-IP). *Journal of Operational Oceanography*, *8*(Suppl. 1), s80–s97. https://doi.org/10.1080/1755876X.2015.1022329

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluations of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1132–1161. https://doi.org/10.1002/qj.2063 [Data available from University of Hamburg ICDC data center (http://icdc.cen.uni-hamburg.de), under Re-Analyses: Ocean]

Bromirski, P. D., Miller, A. J., Flick, R. E., & Auad, G. (2011). Dynamical suppression of sea level rise along the pacific coast of North America: Indications for imminent acceleration. *Journal of Geophysical Research*, *116*, C07005. https://doi.org/10.1029/2010JC006759

Cahoon, D. R. (2015). Estimating relative sea-level rise and submergence potential at a coastal wetland. *Estuaries and Coasts*, *38*, 1077–1084. https://doi.org/10.1007/s12237-014-9872-8

Calafat, F. M., Chambers, D. P., & Tsimplis, M. N. (2014). On the ability of global sea level reconstructions to determine trends and variability. *Journal of Geophysical Research: Oceans*, *119*, 1572–1592. https://doi.org/10.1002/2013JC009298

Carson, M., Köhl, A., & Stammer, D. (2015). The impact of regional multidecadal and century-scale internal climate variability on sea level trends in CMIP5 models. *Journal of Climate*, *28*(2), 853–861. https://doi.org/10.1175/JCLI-D-14-00359.1

Carton, J. A., & Giese, B. S. (2008). A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA). *Monthly Weather Review*, *136*(8), 2999–3017. [Data are available from, e.g., the ERDDAP data server on NOAA's CoastWatch webpage: http://coastwatch.pfeg.noaa.gov/]

Carton, J. A., Giese, B. S., & Grodsky, S. A. (2005). Sea level rise and the warming of the oceans in the Simple Ocean Data Assimilation (SODA) ocean reanalysis. *Journal of Geophysical Research*, *110*, C09006. https://doi.org/10.1029/2004JC002817

Chepurin, G. A., Carton, J. A., & Leuliette, E. (2014). Sea level in ocean reanalyses and tide gauges. *Journal of Geophysical Research: Oceans*, *119*, 147–155. https://doi.org/10.1002/2013JC009365

Church, J. A., & White, N. J. (2011). Sea-level rise from the late 19th to the early 21st century. *Surveys in Geophysics*, *32*, 585–602. https://doi.org/10.1007/s10712-011-9119-1

Church, J. A., White, N. J., Coleman, R., Lambeck, K., & Mitrovica, J. X. (2004). Estimates of the regional distribution of sea level rise over the 1950–2000 period. *Journal of Climate*, *17*, 2609–2625. https://doi.org/10.1175/1520-0442(2004)017<2609:EOTRDO>2.0.CO;2 [Data are available from http://doi.org/10.4225/08/551860B49A2BC]

Dangendorf, S., Marcos, M., Wöppelmann, G., Conrad, C. P., Frederikse, T., & Riva, R. (2017). Reassessment of 20th century global mean sea level rise. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(23), 5946–5951. https://doi.org/10.1073/pnas.1616007114

Fu, L.-L., & Cazenave, A. (2001). *Satellite altimetry and earth sciences, International geophysical series* (Vol. 69). San Diego, CA: Academic Press.

Greatbatch, R. J. (1994). A note on the representation of steric sea level in models that conserve volume rather than mass. *Journal of Geophysical Research*, *99*, 12767–12771.

Griffies, S. M., & Greatbatch, R. J. (2012). Physical processes that impact the evolution of global mean sea level in ocean climate models. *Ocean Modelling*, *51*, 37–72. https://doi.org/10.1016/j.ocemod.2012.04.003

Hamlington, B. D., Leben, R. R., Nerem, R. S., Han, W., & Kim, K.-Y. (2011). Reconstructing sea level using cyclostationary empirical orthogonal functions. *Journal of Geophysical Research*, *116*, C12015. https://doi.org/10.1029/2011JC007529 [Data are available from http://doi.org/10.5067/RECSL-000V1]

Hamlington, B. D., Leben, R. R., Nerem, R. S., & Kim, K.-Y. (2012). Improving sea level reconstructions using non-sea level measurements. *Journal of Geophysical Research*, *117*, C10025. https://doi.org/10.1029/2012JC008277

Han, W., Meehl, G. A., Rajagopalan, B., Fasullo, J. T., Hu, A., Lin, J., . . . Yeager, S. (2010). Patterns of Indian Ocean sea-level change in a warming climate. *Nature Geoscience*, *3*, 546–550. https://doi.org/10.1038/ngeo901

Hay, C. C., Morrow, E., Kopp, R. E., & Mitrovica, J. X. (2015). Probabilistic reanalysis of twentieth-century sea-level rise. *Nature*, *517*, 481–484. https://doi.org/10.1038/nature14093

Hong, B. G., Sturges, W., & Clarke, J. (2000). Sea level on the U.S. east coast: Decadal variability caused by open ocean wind curl forcing. *Journal of Physical Oceanography*, *30*, 2088–2098.

Kaplan, A., Kushnir, Y., Cane, M. A., & Blumenthal, M. B. (1997). Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *Journal of Geophysical Research*, *102*, 27835–27860.

King, M. A., Keshin, M., Whitehouse, P. L., Thomas, I. D., Milne, G., & Riva, R. E. M. (2012). Regional biases in absolute sea-level estimates from tide gauge data due to residual unmodeled vertical land movement. *Geophysical Research Letters*, *39*, L14806. https://doi.org/10.1029/2012GL052348

Köhl, A. (2015). Evaluation of the GECCO2 ocean synthesis: Transports of volume, heat and freshwater in the Atlantic. *Quarterly Journal of the Royal Meteorological Society*, *141*(686), 166–181. [Data available from University of Hamburg ICDC data center (http://icdc.cen.uni-hamburg.de), under Re-Analyses: Ocean]

Köhl, A., & Stammer, D. (2008). Decadal sea level changes in the 50-yr GECCO ocean synthesis. *Journal of Climate*, *21*, 1876–1890. https://doi.org/10.1175/2007JCLI2081.1

Kolker, A. S., Allison, M. A., & Hameed, S. (2011). An evaluation of subsidence rates and sea-level variability in the northern Gulf of Mexico. *Geophysical Research Letters*, *38*, L21404. https://doi.org/10.1029/2011GL049458

Lombard, A., Garric, G., & Penduff, T. (2009). Regional patterns of observed sea level change: Insights from a 1/4° global ocean/sea-ice hindcast. *Ocean Dynamics*, *59*, 433–449. https://doi.org/10.1007/s10236-008-0161-6

Meyssignac, B., Becker, M., Llovel, W., & Cazenave, A. (2012a). An assessment of two-dimensional past sea level reconstructions over 1950–2009 based on tide-gauge data and different input sea level grids. *Surveys in Geophysics*, *33*, 945–972. https://doi.org/10.1007/s10712-011-9171-x [Data are available from ftp://ftp.legos.obsmip.fr/pub/2D_sealevel_reconstruction/]

Meyssignac, B., Salas y Melia, D., Becker, M., Llovel, W., & Cazenave, A. (2012b). Tropical Pacific spatial trend patterns in observed sea level: Internal variability and/or anthropogenic signature? *Climate of the Past*, *8*, 787–802. https://doi.org/10.5194/cp-8-787-2012

Palanisamy, H., Meyssignac, B., Cazenave, A., & Delcroix, T. (2015). Is anthropogenic sea level fingerprint already detectable in the Pacific Ocean? *Environmental Research Letters*, *10*, 084024. https://doi.org/10.1088/1748-9326/10/8/084024

Peltier, W. (2004). Global glacial isostasy and the surface of the ice-age earth. *Annual Review of Earth and Planetary Sciences*, *32*, 111–149.

Permanent Service for Mean Sea Level. (2015). *Tide gauge data*. Retrieved from www.psmsl.org/data/obtaining/

Piecuch, C. G., Dangendorf, S., Ponte, R. M., & Marcos, M. (2016). Annual sea level changes on the North American Northeast Coast: Influence of local winds and barotropic motions. *Journal of Climate*, *29*, 4801–4816. https://doi.org/10.1175/JCLI-D-16-0048.1

Ray, R. D., & Douglas, B. C. (2011). Experiments in reconstructing twentieth-century sea levels. *Progress in Oceanography*, *91*, 496–515. https://doi.org/10.1016/j.pocean.2011.07.021

Roemmich, D., & Owens, W. B. (2000). The Argo Project: Global ocean observations for the understanding and prediction of climate variability. *Oceanography*, *13*(2), 45–50.

Santamaría-Gómez, A., Gravelle, M., & Wöppelman, G. (2014). Long-term vertical land motion from double-differenced tide gauge and satellite altimetry data. *Journal of Geodesy*, *88*, 207–222. https://doi.org/10.1007/s00190-013-0677-5

Serazin, G., Meyssignac, B., Penduff, T., Terray, L., Barnier, B., & Molines, J. (2016). Quantifying uncertainties on regional sea level change induced by multidecadal intrinsic oceanic variability. *Geophysical Research Letters*, *43*, 8151–8159. https://doi.org/10.1002/2016GL069273

Slangen, A. B. A., van de Wal, R. S. W., Wada, Y., & Vermeersen, L. L. A. (2014). Comparing tide gauge observations to regional patterns of sea-level change (1961–2003). *Earth System Dynamics*, *5*, 243–255. https://doi.org/10.5194/esd-5-243-2014

Song, T. Y., & Colberg, F. (2011). Deep ocean warming assessed from altimeters, Gravity Recovery and Climate Experiment, in situ measurements, and a non-Boussinesq ocean general circulation model. *Journal of Geophysical Research*, *116*, C02020. https://doi.org/10.1029/2010JC006601

Stammer, D., Cazenave, A., Ponte, R. M., & Tamisiea, M. E. (2013). Causes for contemporary regional sea level changes. *Annual Review of Marine Science*, *5*, 21–46. https://doi.org/10.1146/annurev-marine-121211-172406

Stammer, D., Köhl, A., Awaji, T., Carton, M. B. D. B. J., Ferry, N., Fischer, A., . . . Xue, Y. (2010). Ocean information provided through ensemble ocean syntheses. In J. Hall, D. E. Harrison, and D. Stammer (Eds.), *Proceedings of the OceanObs'09: Sustained ocean observations and information for society conference* (Vol. 2, ESA Publ. WPP-306). Venice, Italy.

Storto, A., Masina, S., Balmaseda, M., Guinehut, S., Xue, Y., Szekely, T., . . . Wang, X. (2015). Steric sea level variability (1993–2010) in an ensemble of ocean reanalyses and objective analyses. *Climate Dynamics*, *49*, 709–729. https://doi.org/10.1007/s00382-015-2554-9

Sturges, W., & Douglas, B. C. (2011). Wind effects on estimates of sea level rise. *Journal of Geophysical Research*, *116*, C06008. https://doi.org/10.1029/2010JC006492

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, *305*(5683), 503–505. https://doi.org/10.1126/science.1099192

Vinogradov, S. V., & Ponte, R. M. (2011). Low-frequency variability in coastal sea level from tide gauges and altimetry. *Journal of Geophysical Research*, *116*, C07006. https://doi.org/10.1029/2011JC007034

Watson, C. S., White, N. J., Church, J. A., King, M. A., Burgette, R. J., & Legresy, B. (2015). Unabated global mean sea-level rise over the satellite altimeter era. *Nature Climate Change*, *5*, 565–569. https://doi.org/10.1038/nclimate2635

Wenzel, M., & Schröter, J. (2010). Reconstruction of regional mean sea level anomalies from tide gauges using neural networks. *Journal of Geophysical Research*, *115*, C08013. https://doi.org/10.1029/2009JC005630

Wenzel, M., & Schröter, J. (2014). Global and regional sea level change during the 20th century. *Journal of Geophysical Research: Oceans*, *119*, 7493–7508. https://doi.org/10.1002/2014JC009900 [Data are available from ftp://ftp.legos.obsmip.fr/pub/2D_sealevel_reconstruction/]

White, N. J., Church, J. A., & Gregory, J. M. (2005). Coastal and global averaged sea level rise for 1950 to 2000. *Geophysical Research Letters*, *32*, L01601. https://doi.org/10.1029/2004GL021391

Wöppelmann, G., & Marcos, M. (2016). Vertical land motion as a key to understanding sea level change and variability. *Reviews of Geophysics*, *54*, 64–92. https://doi.org/10.1002/2015RG000502

Wunsch, C., & Heimbach, P. (2007). Practical global oceanic state estimation. *Physica D*, *230*, 197–208. https://doi.org/10.1016/j.physd.2006.09.040

Wunsch, C., & Stammer, D. (1997). Atmospheric loading and the oceanic "inverted barometer" effect. *Reviews of Geophysics*, *35*, 79–107.

Zhang, X., & Church, J. A. (2012). Sea level trends, interannual and decadal variability in the Pacific Ocean. *Geophysical Research Letters*, *39*, L21701. https://doi.org/10.1029/2012GL053240