

Key Points:

- Several data sets and noise models are used to compute global and regional steric sea-level change for 1993–2017 and 2005–2015
- Global mean steric sea-level trend estimates differ up to 2 mm/year depending on the data set and noise model used
- Regional sea-level trends require more complex noise models and exhibit spatial coherency in the noise model preference

Supporting Information:

- Supporting Information S1

Correspondence to:

C. M. L. Camargo,
carolina.camargo@nioz.nl

Citation:



Camargo, C. M. L., Riva, R. E. M., Hermans, T. H. J., & Slangen, A. B. A. (2020). Exploring sources of uncertainty in steric sea-level change estimates. *Journal of Geophysical Research: Oceans*, 125, e2020JC016551. <https://doi.org/10.1029/2020JC016551>

Received 26 JUN 2020

Accepted 21 SEP 2020

Accepted article online 25 SEP 2020

Exploring Sources of Uncertainty in Steric Sea-Level Change Estimates

Carolina M. L. Camargo^{1,2} , Riccardo E. M. Riva² , Tim H. J. Hermans^{1,2} , and Aimée B. A. Slangen¹ 

¹Department of Estuarine and Delta Systems, NIOZ Royal Netherlands Institute for Sea Research and Utrecht University, Yerseke, Netherlands, ²Department of Geoscience and Remote Sensing, Delft University of Technology, Delft, Netherlands

Abstract Recent studies disagree about the contribution of variations in temperature and salinity of the oceans—steric change—to the observed sea-level change. This article explores two sources of uncertainty to both global mean and regional steric sea-level trends. First, we analyze the influence of different temperature and salinity data sets on the estimated steric sea-level change. Next, we investigate the impact of different stochastic noise models on the estimation of trends and their uncertainties. By varying both the data sets and noise models, the global mean steric sea-level trend and uncertainty can vary from 0.69 to 2.40 and 0.02 to 1.56 mm/year, respectively, for 1993–2017. This range is even larger on regional scales, reaching up to 30 mm/year. Our results show that a first-order autoregressive model is the most appropriate choice to describe the residual behavior of the ensemble mean of all data sets for the global mean steric sea-level change over the last 25 years, which consequently leads to the most representative uncertainty. Using the ensemble mean and the first-order autoregressive noise model, we find a global mean steric sea-level change of 1.36 ± 0.10 mm/year for 1993–2017 and 1.08 ± 0.07 mm/year for 2005–2015. Regionally, a combination of different noise models is the best descriptor of the steric sea-level change and its uncertainty. The spatial coherence in the noise model preference indicates clusters that may be best suited to investigate the regional sea-level budget.

Plain Language Summary Ocean temperature and salinity variations lead to changes in sea level, known as steric sea-level change. Steric variations are important contributors to sea-level change and reflect how the oceans have been responding to global warming. For this reason, several recent studies have quantified the contribution of steric variations to global and regional sea-level change. However, the reported rates largely differ between studies. In this paper, we look at how the use of different temperature and salinity data sets can be one of the causes of the different estimates of steric sea-level change published so far. We also investigate how different methods (noise models) used to obtain the rate of change can be another source of different results. We find that the rate of change can vary up to 2 mm/year for the global mean as a result of different data sets and methods used. Regionally, differences can reach up to several tens of millimeters per year. We show that the noise models should always be carefully chosen for each region, so that the rate of change is accurately estimated.

1. Introduction

The oceans are a major reservoir of heat and have stored about 90% of the human-induced heat in the climate system over the last 50 years (MacIntosh et al., 2017; von Schuckmann et al., 2016). The resulting ocean warming leads to sea-level change (SLC), which is an important reflection of how the oceans respond to global warming (Kopp et al., 2016). This process, known as thermosteric SLC, is the dominant component of the steric contribution to global mean sea-level (GMSL) change.

Steric SLC is also partially driven by salinity variations (i.e., halosteric change). The contribution of salinity to GMSL change is negligible, as the ocean's salt content is considered to be constant over multidecadal time scales (Cazenave et al., 2018). Regionally, however, halosteric SLC can be just as important as thermosteric changes, or even the dominant process of steric sea-level variations, for instance, in polar regions (Stammer et al., 2013). Hence, studies of steric contributions to regional SLC should consider both salinity and temperature variations (MacIntosh et al., 2017).

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

In addition to density-driven variations, present-day SLC is also driven by ocean mass fluctuations (Church et al., 2013). The main processes that change the amount of mass available in the oceans are the melting of glaciers and ice sheets and variations in terrestrial water storage. Together, density and mass variations represent the total budget of SLC (e.g., Cazenave et al., 2018; Gregory et al., 2013). Thus, accurately quantifying steric sea-level variations and their uncertainties constrains the other contributions to the observed SLC.

Comparing published rates of steric GMSL change is complicated, as the period considered differs between studies. For example, while the fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR4, Bindoff et al., 2007) estimated a thermosteric contribution of 1.6 ± 0.5 mm/year to the GMSL change for 1993–2003, the IPCC AR5 (Church et al., 2013) reported a thermosteric contribution of 1.1 ± 0.3 mm/year for 1993–2010. In addition to the different study periods, part of the difference between steric rates can also be explained by measurement biases, such as depth biases in expandable and mechanical bathythermograph (XBTs and MBTs) observations (Ishii and Kimoto, 2009). However, even if one considers the same time period and the same corrections, estimates of steric SLC can still differ for the following three reasons.

First, the deep ocean (>2,000 m depth) contribution to steric SLC is uncertain. Given the lack of observational data below 2,000 m depth, different estimates of how much the deep ocean has contributed to SLC have been made in the recent past (e.g., Desbruyères et al., 2016; Dieng et al., 2015; Llovel et al., 2014; Purkey and Johnson, 2010). For instance, based on the few high-quality full-depth hydrographic measurements available, Purkey and Johnson (2010) estimated that deep ocean warming contributed 0.1 mm/year to the GMSL change from 1980 to 2010. Contrarily, through a budget analysis, Llovel et al. (2014) found a negative (-0.13 ± 0.34 mm/year) deep ocean contribution to GMSL change from 2005 to 2013. According to an ocean model simulation, repeated hydrographic measurements tend to underestimate the deep ocean warming due to biases induced by temporal and spatial sampling (Garry et al., 2019).

A second source of uncertainty in steric SLC is the use of different temperature and salinity data and how these data are processed. Dieng et al. (2015) found that the steric sea-level trend varied from 0.15 to 0.92 mm/year, from 2003 to 2012, depending on the data set analyzed. These differences can be explained mainly by the corrections made to individual measurements and by different data processing methodologies, such as gap filling methods (Dieng et al., 2015). Furthermore, the type of the data used (i.e., in situ observations, reanalysis (REA), or ocean model output) can lead to diverging trends.

Finally, the method used to determine the trend from a time series is another source of uncertainty. Trends are usually obtained by fitting a linear or quadratic model to a time series. The associated error can be described by different stochastic noise models (Bos et al., 2013). Bos, Williams, et al. (2014) showed that uncertainties in SLC can be underestimated when an improper noise model is used. While the effect of using different noise models on estimates of the total SLC from satellite altimetry and tide gauges has already been discussed (e.g., Bos, Williams, et al., 2014; Hughes and Williams, 2010; Royston et al., 2018), its effect on estimates of steric SLC is still unclear. Given that the choice of the noise model is highly dependent on the nature of the process it describes (Royston et al., 2018), here we investigate the impact of different noise models on steric sea-level trends.

This study explores how steric sea-level trends can vary, based on the data sets and the noise models used. We limit our analysis to the upper ocean (0–2,000 m), where most data are available, and leave studying deep ocean uncertainties for future work. We compute and compare steric sea-level trends from 1993 to 2017 (satellite altimetry era) and from 2005 to 2015 (Argo period), considering 15 different gridded temperature and salinity data sets and 8 different noise models. We will show that the uncertainties of both data sets and noise models play an important role in estimating the steric contribution to SLC.

We describe how we calculated steric SLC and the ocean temperature and salinity data sets used in section 2, followed by an explanation of the noise models used to compute the trends and uncertainties. We show the results of the different data sets and noise models for the global mean and regional steric SLC in sections 3 and 4, respectively. We display mainly the results for 2005–2015, when it was inconvenient to display both periods in the same figures. The complementary figures for 1993–2017 can be found in the supplementary information.

Table 1
Temperature and Salinity Data Sets Used to Compute Steric SLA, Separated Into Three Categories

Category	Abbreviation	Version	Institute	Reference	Time period	Depth range (m)
Argo	APDRC		IPRC/APDRC	Product Description	2005–2019.06	0–2,000
	SIO		Scripps Institution of Oceanography	Roemmich and Gilson (2009)	2004–2019.06	10–1,975
	BOA		CSIO	Li et al. (2017) and Lu et al. (2019)	2004–2018	0–1,975
Multiple in situ (MiS)	ISAS	15	Ifremer	Gaillard et al. (2016)	2002–2015	0–2,000
	CORA	v5.2	Ifremer	Cabanes et al. (2013) and Szekely et al. (2019)	1940–2017	0–2,000
	EN4	v4.2G10 ^a	UK Met Office	Good et al. (2013)	1950–2018	5–5,350
	IK09	v7	JAMSTEC	Ishii and Kimoto (2009) and Ishii et al. (2017)	1900–2019.05	5–3,000
	C17	v0 (S), v3(T)	IAP/CAS	Cheng et al. (2017)	1940–2019.05	0–2,000
Reanalysis (REA)	ISAS+	15	Ifremer	Gaillard et al. (2016)	2002–2015	0–2,000
	ARMOR	3D	CSL	Guinehut et al. (2012)	1955–2018	0–5,500
	C-GLORSa ^b	05	CMCC	Storto and Masina (2016)	1993–2018	0–5,902
	GLORYS2a ^b	v4	Mercator Ocean	Garric and Parent (2017)	1993–2017	0–5,902
	FOAM-GloSea5a ^b (hereafter FOAM)	v13	UK Met Office	Blockley et al. (2014) and Maclachlan et al. (2015)	1993–2017	0–5,902
	ORASa ^b	5	ECMWF	Zuo et al. (2019)	1993–2017	0–5,902
	SODA	3	UMD/ATMOS	Carton et al. (2018)	1980–2017	0–5,902

Note. The abbreviations are short forms for either the reference paper (e.g., C17 for Cheng et al., 2017), the responsible institution (e.g., APDRC), or the official name of the data set (e.g., CORA). Time period in bold indicates the data sets available for the entire satellite altimetry era.

^aTwo versions of EN4v4.2 are available, depending on the type of correction applied for the XBT and MBT data. Following the description in the discussion of Cheng et al. (2016) about the benefits of each correction, we use the data set with the Gouretski and Reseghetti (2010) correction. ^bThese products were obtained via the Global Ocean Ensemble Reanalysis Version 2 of Mercator Ocean (GLOBAL_REANALYSIS_PHY_001_031).

2. Methods and Data

2.1. Calculating Steric Sea-Level Anomalies

Following Gill and Niller (1973) and Tomczak and Godfrey (2003), the steric sea-level anomaly (SLA) η_s can be described as follows:

$$\eta_s = -\frac{1}{\rho_0} \int_{z_2}^{z_1} \rho' dz, \quad (1)$$

where ρ_0 is a reference density and ρ' is the local density anomaly. The expression is vertically integrated from the maximum local depth z_2 to the water surface z_1 . We use the Thermodynamic Equation of Seawater (TEOS-10, McDougall and Barker, 2011) as the equation of state to calculate the steric SLA and consider only the upper ocean (0–2,000 m) in our computations.

The density anomaly ρ' is a function of variations in the ocean temperature, salinity, and pressure fields and is computed using ocean temperature and salinity data from in situ and REA data sets (section 2.2).

2.2. Ocean Temperature and Salinity Data Sets

We collected and analyzed 15 publicly available gridded data sets of continuously monthly ocean temperature and salinity published by different research groups worldwide (Table 1). All data sets were downloaded between January and March 2020 and are currently accessible (last check June 2020). The data sets were selected based on their recurrent use in the literature. Given that we use gridded data sets, we do not consider the uncertainty propagation from a single profile of temperature and salinity to steric height. The concept of uncertainty propagation was reviewed and discussed in detail in MacIntosh et al. (2017).

The data sets were categorized according to their data type: in situ measurements versus ocean REA. The first category includes ship-based hydrographic data (such as conductivity-temperature-depth profiles, XBTs, and MBTs) and observations collected by Argo profiling floats (Roemmich et al., 2009). The in situ

observations are generally merged onto a grid using statistical interpolation (Storto et al., 2017). This category is further divided into two categories: Argo, for data sets that have only data from Argo floats, and multiple in situ (MiS), for products that combine several sources of in situ observations, in addition to Argo data. The second category, ocean REA, assimilates the observational data into ocean circulation models, extrapolating the information spatially and, sometimes, temporally (MacIntosh et al., 2017).

2.3. Data Processing

First, the steric SLA was computed on the native grid of each data set following Equation 1. We then standardized the varying resolution by remapping all data sets on a 1° by 1° grid, using bilinear interpolation. Next, we selected the grid points within 66°S to 66°N of latitude and applied a land mask based on ETOPO1 (Amante and Eakins, 2009). The land mask was applied without extrapolating any values, meaning that areas which are not covered by all data sets are still considered, using the available data sets in each location. Figure S2 shows the number of data sets available at each grid cell after applying the land mask. Using a more conservative ocean mask, in which only the grid points that all data sets have data are considered, leads to a minor difference of 0.08 mm/year for the global mean trend. Hence, we decided to use the ETOPO1 mask, which has the advantage of a wider spatial coverage. Next, we computed a category mean for each of the three categories (Argo, MiS, and REA) and a total ensemble mean of all individual data sets. Using an area-weighted mean, we computed the global mean steric SLC for each data set.

2.4. Estimating Trends and Uncertainties

The trends and respective uncertainties were estimated using the Hector software (Bos et al., 2013). Hector uses a weighted least squares model to estimate the linear trend, while accounting for periodic signals (i.e., seasonal and annual components). The variations not captured by the regression model are defined as noise (Bos, Williams, et al., 2014). The software allows the user to choose between a number of stochastic models to describe the noise properties, such as temporal correlation and spectral density behavior, thus reducing the risk of underestimating the trend (Bos et al., 2013). The uncertainties provided by Hector and shown in this paper represent one standard deviation.

Based on sea-level literature (e.g., Bos, Williams, et al., 2014; Royston et al., 2018), we tested eight models to find the best descriptor of the uncertainties in our data: a white noise (WN) model; a generalized Gauss-Markov (GGM) model; a pure power law (PL) model; a PL model combined with WN (PLWN); autoregressive models of orders 1, 5, and 9 (AR(1), AR(5), and AR(9), respectively); and an autoregressive fractionally integrated moving average model of order 1 (ARFIMA, henceforth ARF). The WN model is the simplest stochastic model, in which the spectral density of the noise is equal to zero, and no temporal correlation between the residuals is considered. In the PL model, all observations influence one another, although their correlation decreases with increasing temporal distance. The AR(p) models consider that each observation at time t_i is affected by white noise and the observation at time t_{i-1} . The time of preceding observations influencing each observation is given by the order p of the autoregressive model. To represent PL observations at low frequencies, the ARF model combines an AR(1) model with a fractional integration and a moving average of the noise. The GGM model is a generalized form of the ARF model. More details about each model can be found in the Hector software manual and in Bos et al. (2013), Bos, Fernandes, et al. (2014), and Bos, Williams, et al. (2014).

To determine the goodness of fit of the noise models, we use the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978), together with visual inspection of the power spectrum of the noise model and the fit residual. Both AIC and BIC use the maximum likelihood and the number of parameters to judge the quality of the model. However, BIC penalizes the number of parameters stronger than does AIC (Liddle, 2004). AIC and BIC are the most common criteria used to describe the goodness of the fit, yet there is no consensus in the literature about which criterion should be used. Therefore, we used these criteria to select the “best” noise model for each data set and then computed an overall ranking of each noise model. Following Royston et al. (2018), we consider the mean scoring of AIC and BIC to decide which noise model is the most suitable. More details about the selection procedure is given in supporting information Text S1.

We complemented the global mean noise model analysis with the software ARMASA (Broersen, 2020). Given a residual time series, ARMASA identifies the ideal noise model type and order, choosing between

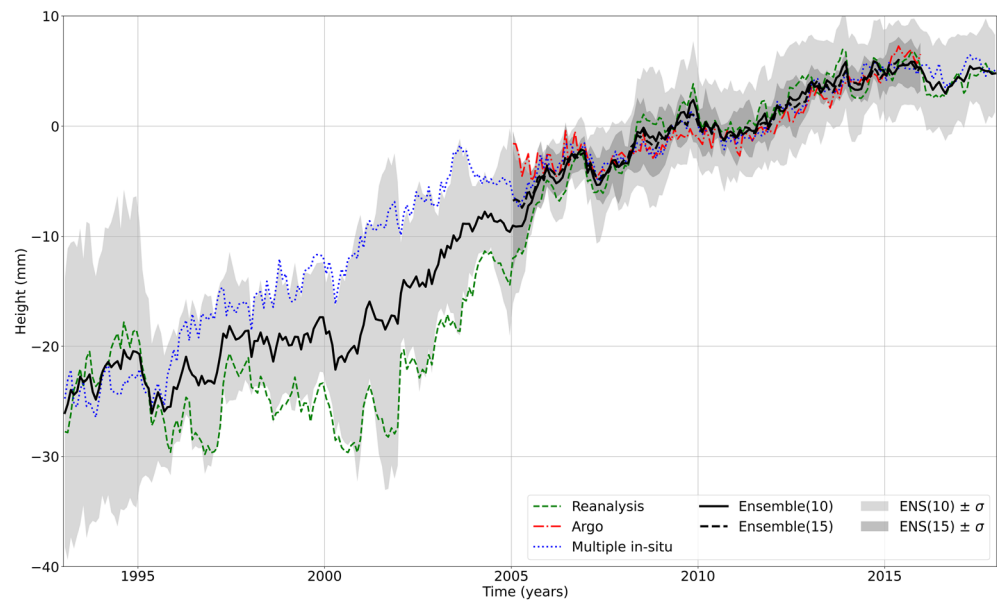


Figure 1. Global mean steric sea-level time series of the ensemble (black solid and dashed lines) and category means (green for reanalysis, red for only Argo, and blue for multiple in situ) referenced to the time mean for 2005–2015. The solid black line and the light gray shaded area represent the ensemble of 10 data sets for 1993–2017 ± 1 standard deviation, respectively. The dashed black line and the dark gray shaded area represent the ensemble of 15 data sets for 2005–2015 ± 1 standard deviation, respectively.

different autoregressive, moving average, and autoregressive moving average models ($AR(p)$, $MA(q)$, and $ARMA(p', p' - 1)$, respectively). More details about ARMASA and the parameters used are given in supporting information Text S2 and Klees and Broersen (2002).

3. Global Mean Steric SLC

3.1. Steric Sea-Level Time Series

From the temperature and salinity data sets (Table 1), we computed 15 global mean steric sea-level time series, three category means (red, blue, and green lines for Argo, MiS, and REA, respectively), and one ensemble mean (black line) (Figure 1). Despite the differences among the time series, all the category means agree within one standard deviation from the ensemble mean. Due to the addition of Argo data, starting in 2002 and reaching a near-global coverage in 2005, the standard deviation of the global mean steric sea level reduces strongly: from about 8 mm/year during the first 10 years to 4.2 mm/year from 2002 to 2005 and to 3.6 mm/year after 2005. This reduction is seen not only for the ensemble mean and Argo data sets but also for the MiS and REA categories, suggesting that all data sets strongly depend on the Argo float data.

The ensemble means based on the 10 and 15 data sets (for 1993–2017 and 2005–2015, in Figure 1, solid and dashed black lines, respectively) behave very similarly for the overlapping period. This indicates the robustness of the ensemble mean to the number of data sets used. Furthermore, the category means agree well with the ensemble mean for 2005–2015, all within one standard deviation from the ensemble.

3.2. Influence of Data Sets and Noise Models on Steric Estimates

For each of the time series, category, and ensemble means, we computed linear trends using eight different noise models, leading to 104 (13×8) and 152 (19×8) global mean trends for 1993–2017 and 2005–2015, respectively (Figure 2 for 2005–2015 and Figure S3 for 1993–2017). When both the data set and noise model are varied simultaneously for the 2005–2015 period ($n = 15 \times 8 = 120$), the global mean trend (Figure 2a) ranges from 0.56 to 2.33 mm/year. Likewise, the uncertainty (Figure 2b) ranges from 0.02 to 1.65 mm/year.

The situation is similar when a single noise model is chosen and the data sets are varied. For example, in the case of $AR(1)$, the spread of the trend and uncertainty, although reduced, continues to be large, respectively

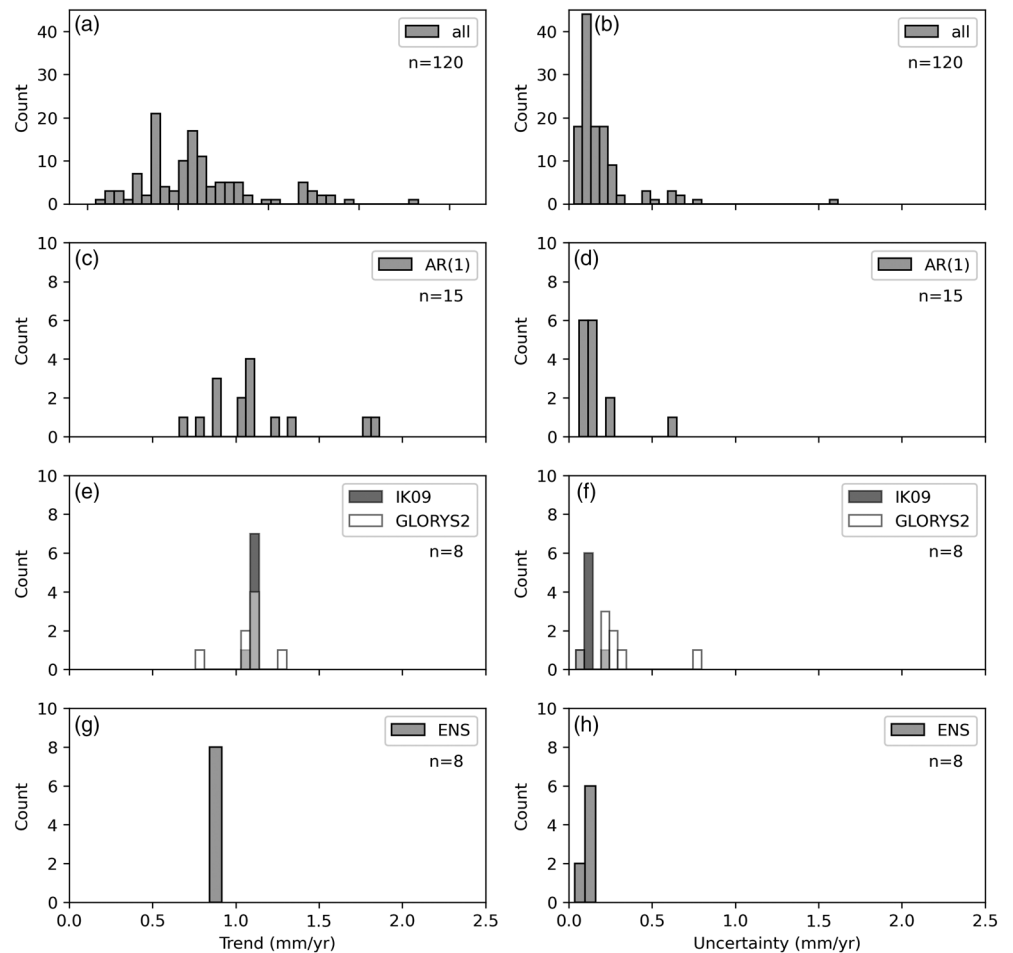


Figure 2. Histogram of the (left column) trend and (right column) uncertainty in millimeters per year for 2005–2015 (a, b) varying all data sets and noise models, (c, d) for the AR(1) noise model when varying all the data sets, (e, f) for the IK09 (white) and GLORYS2 (gray) data sets when varying all the noise models, and (g, h) for the ensemble mean when varying all the noise models. The count on the y-axis is with respect to the number of data sets (n) considered in each case.

ranging from 0.66 to 1.85 and 0.05 to 0.65 mm/year (Figures 2c and 2d, $n = 15$). Alternatively, when we choose one data set and vary only the noise model (Figures 2e–2h, $n = 8$), the spread of the trend becomes significantly smaller. This suggests that varying the data set has a larger influence on the estimated trend than has varying the noise model. In addition, we see that the spread is highly dependent on the chosen data set (Figures 2e and 2f). While IK09 (gray) has a minimal spread as a result of varying the noise models, GLORYS2 (white) shows a variation of almost 1 mm/year depending on the noise model being used. Interestingly, the ensemble mean trend and uncertainty (Figures 2g and 2h) show almost no sensitivity to the choice of noise model. This could be the result of the ensemble mean being smoother than the individual time series, and it indicates that using the ensemble mean reduces the dependency on the noise model choice. The results are similar for the period 1993–2017 (Figure S3).

3.3. Noise Model Selection (Global Mean)

In order to determine which noise model best describes the global mean steric trends and uncertainties, we use the AIC and BIC scores (section 2.4). Our selection criterion is based on a threshold value (see Text S1 for more information); thus, multiple noise models can be selected for each data set if the score passes the threshold, and the final percentage of each score can be larger than 100.

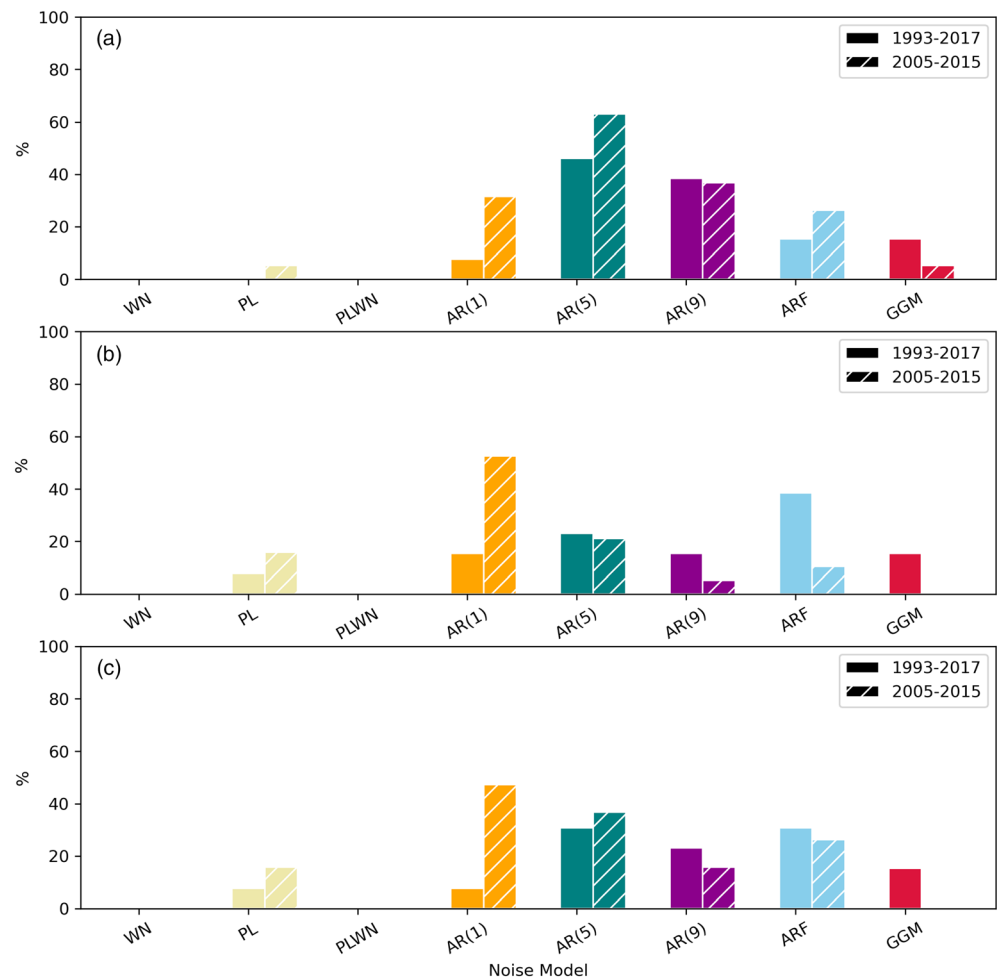


Figure 3. Percentage of preferred noise model based on all data sets (including the ensemble and category means) for 2005–2015 (hashed bars) and 1993–2017 (full bars) according to (a) AIC, (b) BIC, and (c) the mean of AIC and BIC.

According to AIC, AR(5) is the overall preferred noise model (78% for 2005–2015, Figure 3, hashed bars; and 56% for 1993–2017, Figure 3, full bars). Contrarily, BIC ranks AR(1) first for 2005–2015 and ARF for 1993–2017. When the category and ensemble means are excluded from the selection (not shown), AR(1) is never chosen as the preferred noise model for 1993–2017. This illustrates how the averaging process in ensemble means reduces the temporal variability of the time series, especially when a longer period is considered. Thus, a simple noise model, such as AR(1), becomes the preferred model for the category and ensemble means. For the mean of AIC and BIC, both AR(5) and ARF are equally preferred for 1993–2017, and AR(1) is ranked first for 2005–2015.

We complement our noise model selection with ARMASA (section 2.4). Between $AR(p)$, $MA(q)$, and $ARMA(p', p' - 1)$ models, a simple moving average of order 6 is selected as the preferred noise model for the ensemble mean of 2005–2015. This shows that there is still some short-term periodicity in the residuals, even after the annual and semiannual signals have been removed. Between $AR(p)$ and $ARMA(p', p' - 1)$, the AR(1) is selected as the best noise model for both periods. However, for the individual time series, many different $AR(p)$, $MA(q)$, and $ARMA(p', p' - 1)$ models were chosen (see Table S2). For the individual data sets, the most recurrent noise model for both periods was $ARMA(2, 1)$. The preferred order of the noise models tends to be higher for the time series for 1993–2017 than for 2005–2015. Interestingly, while AIC and BIC found AR(5) to be one of the best noise models, this order was never preferred by ARMASA. This may be related to the much wider range of orders and model types tested by the ARMASA algorithm in order to select the one model that is statistically representative of the data (Broersen, 2002). We have used the

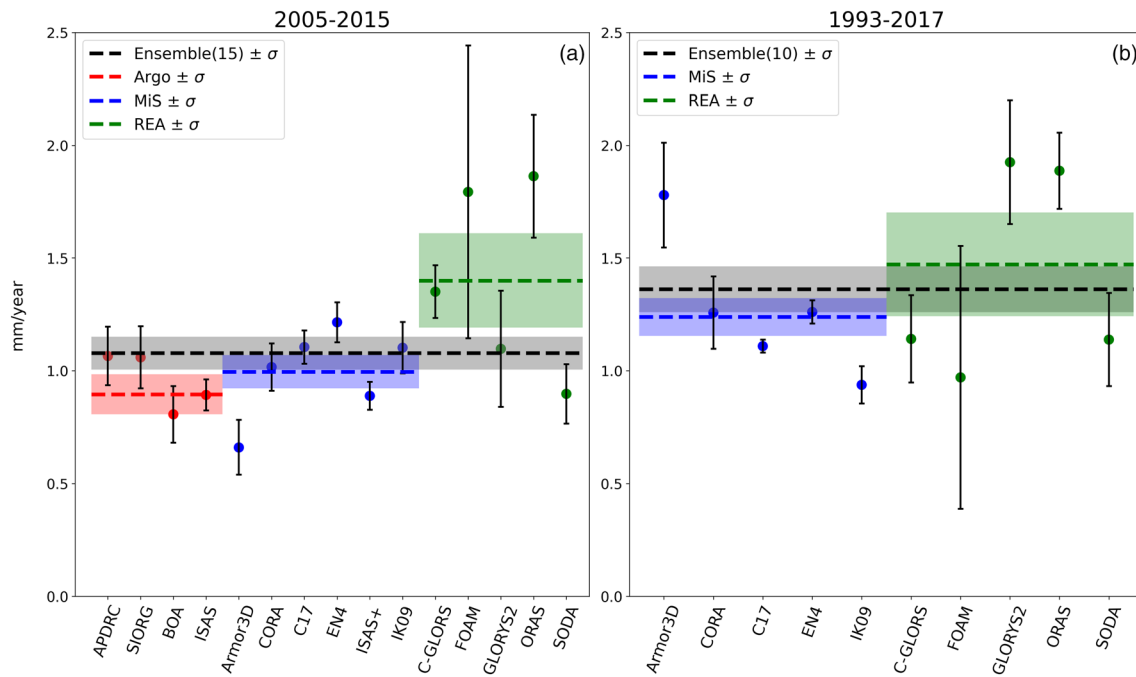


Figure 4. Global mean sea-level trends (mm/year) and uncertainty (1-sigma) obtained with the AR(1) noise model from (a) 2005 to 2015 and (b) 1993 to 2017. Dashed lines represent the trends of the ensemble and category means, with the respective uncertainties in the shaded area. Gray, red, blue, and green indicate the ensemble mean and the Argo, multiple in situ, and reanalysis category means. The means refer to values obtained from the mean of the individual time series, and not the mean of the trends and uncertainties. The trends and uncertainties are also listed in Table S1.

ARMASA results as a qualitative indicator of the ideal noise model and find that while there are differences with the chosen model in Hector, these differences are small.

Considering the AIC and BIC ranking and the results of ARMASA, we concluded that an AR(1) model is a good descriptor of the trend and uncertainties of the global mean time series, especially for the category and ensemble means. Nevertheless, there is still a periodicity in the residuals. We see that for the Argo period, a signal of several months (2 to 6) is still present in the noise, while for the satellite altimetry era, an interannual signal, in the order of 3 up to 20 months, is seen in the residuals.

3.4. Global Mean Steric Sea-Level Trends

Applying an AR(1) noise model, we find a global mean steric sea-level trend of 1.08 ± 0.07 mm/year for the ensemble mean for 2005–2015 (Figure 4a) and 1.36 ± 0.10 mm/year for 1993–2017 (Figure 4b). For most of the data sets, the uncertainties are larger for 1993–2017, mainly as a result of the increased observational quality and coverage in 2005–2015. Overall, there is a better agreement between the observational data sets (i.e., Argo [red dots] and MiS [blue dots]) than between the REA products (green dots).

The Argo category (Figure 4, red) shows a good intracategory agreement with all individual values agreeing within the category uncertainties. The Argo category mean trend of 0.90 ± 0.09 mm/year is in the lower uncertainty range of the APDRC and SIORG data sets. The MiS category (Figure 4, blue) shows large differences for the two time periods, with a category mean of 1.00 ± 0.07 and 1.24 ± 0.08 mm/year for the Argo period and satellite altimetry era, respectively. For 2005–2015, the category mean is within 1-sigma of all individual data sets, except for Armor3D on the lower limit and EN4 on the upper limit of the category distribution. For the 1993–2017 period, the category mean is only within the uncertainty ranges of the CORA and EN4 data sets, with IK09 on the lower limit and Armor3D on the upper limit of the category distribution.

The REA category (Figure 4, green) shows the largest ranges for both periods, with almost 1 mm/year difference between the trends of individual data sets. The category mean for 2005–2015 is 1.40 ± 0.21 mm/year, falling within the uncertainty range of C-GLORS, FOAM, and GLORYS2. The category mean for 1993–2017 is 1.47 ± 0.23 mm/year, within the uncertainty range of all individual trends. The uncertainties

from the REA data sets are much larger than for the other groups, probably related to the higher spatial resolution of such products and the modeled internal variability.

To test if the variance in the ensemble and category means explain the variance of the individual data sets, we computed the R^2 (Table S1). All data sets have a relatively high R^2 value (>0.7) in relation to the ensemble and category means, with exception of the FOAM REA ($R^2 = 0.52$ for 2005–2015 and $R^2 = 0.37$ for 1993–2017), IK09 for 1993–2017 ($R^2 = 0.66$), and the BOA data set for 2005–2015 ($R^2 = 0.68$). Thus, the ensemble mean is representative of almost all the individual members. Furthermore, to test the sensitivity of the ensemble mean, we removed up to two data sets (one of the largest outliers [FOAM and Armor3D] plus any other data set) and recomputed ensemble mean. This led to a maximum variation of the trend of 0.12 mm/year for 2005–2015 and 0.21 mm/year for 1993–2015.

4. Regional Steric SLC

Several processes that are important at a regional scale are averaged out in the GMSL analysis, especially when discussing residuals and uncertainties. Thus, in this section we complement the global mean discussion with results about the regional steric sea-level trends. To show how the steric SLC is regionally influenced by the chosen data sets (section 4.1) and the noise models (section 4.2), we selected one latitudinal transect in the center of each ocean basin (Figure 5a). We then use the ensemble mean to present the regional selection of noise models (section 4.3) used to obtain the regional steric sea-level trends (section 4.4).

4.1. Trend Dependence on Chosen Steric Data Sets

We first compare the steric trends of all different data sets for three latitudinal transects, using the AR(1) noise model, based on the noise model selection analysis in section 3. The ensemble mean (right profile of each transect) represents features of all the other data sets, with smoother transitions. The profiles for 1993–2017 (Figure S4) are dominated by positive trends in all transects and lower uncertainties than are those for 2005–2015, as the effect of interannual variability is reduced and the positive trend becomes clearer.

The Indian Ocean (Figures 5c and 5d) is dominated by a positive steric trend, which can be related to the increasing atmospheric temperature in the region (Carvalho & Wang, 2019). Around 10°S, the trends in all data sets are not statistically significant and reverse to -2 mm/year, accompanied by an increased uncertainty (~ 3 mm/year). Going south, the trends reverse back to positive, reaching the highest values of trend and uncertainty around 40°S (up to 30 ± 10 mm/year for FOAM). Further south (around 55°S) the trend becomes negative again, possibly the effect of negative trends of the Antarctic Circumpolar Current (Frankcombe et al., 2013).

In the Pacific transect (Figures 5e and 5f), the band between 40°S and 60°S shows most variation between data sets, with a striping pattern that varies in intensity and width from data set to data set. A similar behavior was reported in a climate model study by Slangen et al. (2015), who related these changes to aerosol and greenhouse gas forcing. In the southern South Pacific, all data sets have a negative trend. Similarly to the Indian Ocean, the uncertainties in the Pacific follow the pattern of the trends and increase where the trends are reversing. There is a band of high uncertainty around the equator, coinciding with the negative trend values.

Most of the Atlantic Ocean transect trends and uncertainties (Figures 5g and 5h) agree between the data sets, except for the APDRC results. Most of the data sets have a negative trend in the north, matching the mid-2000s' signal of the North Atlantic subpolar gyre (Chafik et al., 2019). However, the length and intensity of this negative band vary significantly between the data sets. Only the CORA data set shows a positive trend in the north. For most of the data sets, around 20°S the trend becomes positive, with a stripe pattern of narrow bands until the far south of the transect. Only for the SIORG data set do we see an overall positive trend from 10°S to 60°S. The uncertainty values are generally lower in the Atlantic compared to those in the other transects. Most of the data sets, especially EN4, show an increase in the uncertainty around 40°S. The atypical negative trend between 40°N and 20°N of the APDRC data set is accompanied by a very high uncertainty value.

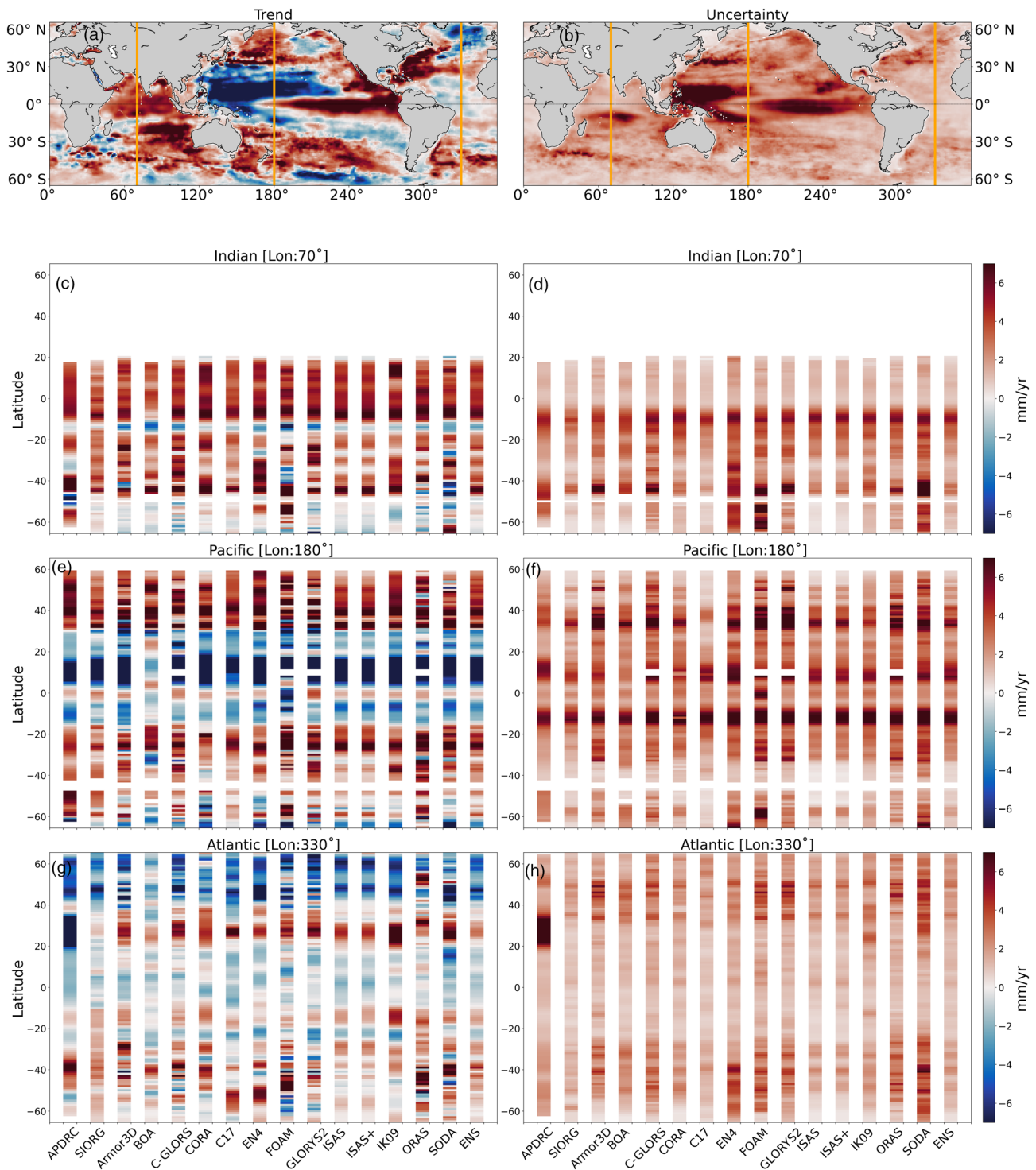


Figure 5. (a) Trend and (b) uncertainty for 2005–2015 estimated with the AR(1) noise model, and the latitudinal transects (displayed in orange), as a function of the data set (x-axis) for the (c, d) Indian, (e, f) Pacific, and (g, h) Atlantic Ocean transects.

4.2. Trend Dependence on Chosen Noise Models

In contrast with varying the data set, there is almost no latitudinal variation in the trend of the ensemble mean when only varying the noise model (Figure 6 for 2005–2015 and Figure S5 for 1993–2017). However, the effect of the different noise models becomes much clearer in the uncertainty profiles. We

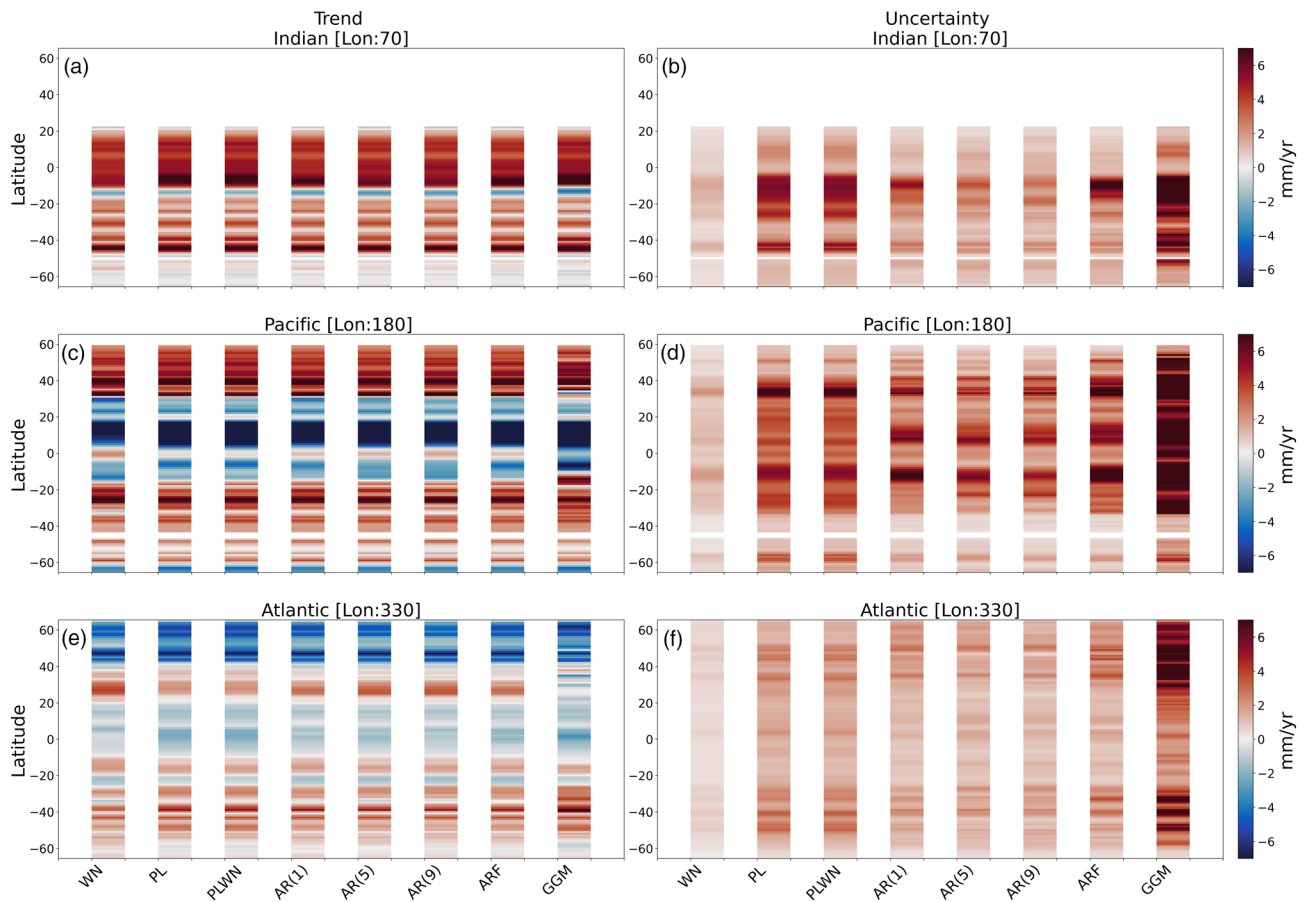


Figure 6. (left column) Trend and (right column) uncertainty for the ensemble mean for 2005–2015, as a function of latitude (y-axis) and data set (x-axis) for the (a, b) Indian, (c, d) Pacific, and (e, f) Atlantic Ocean transects.

see that the WN model is likely underestimating the real uncertainty, because of the low-frequency variability in the time series. Due to the small impact of WN, PL and PLWN show the same behavior. The noise models of the autoregressive family are slightly different from each other, with the uncertainties for AR(1) generally larger than those for AR(5), which in turn are larger than the AR(9) uncertainties. The ARF model has a more pronounced spatial pattern compared to the simple autoregressive models. The GGM model stands out due to very large uncertainties. For 1993–2017 (Figure S5), the latitudinal patterns of the trends and uncertainties are, in general, the same as for 2005–2015.

Figure 6 also illustrates the importance of choosing an appropriate noise model to describe the uncertainties present in the data. The mean uncertainty for the WN model, in all transects, is of the order of ± 0.7 mm/year, while the mean uncertainty for the GGM model is of the order of ± 7 mm/year. However, a very low or high uncertainty does not mean that a specific noise model cannot be the best descriptor of the data. In the next section, we look at the AIC and BIC to find the preferred noise model for each grid point.

4.3. Noise Model Selection (Regional)

To find which noise model best describes the regional variations of the ensemble mean, we investigate the noise model performance at each ocean grid point. These results depend on the resolution and ocean mask used. The regional selection of the preferred noise model (Figure 7 for the ensemble mean and Figure S6 based on individual data sets), shows that AR(1) and AR(5) are the preferred noise models, agreeing with the global analysis (section 3.3). According to AIC (Figures 7a and 7b), AR(5) is the preferred noise model for 41% of the ocean area for 2005–2015 and 53% of the regions for 1993–2017. In contrast, the preference of BIC (Figures 7c and 7d) for AR(1) is very clear, selecting it as the best noise model for 73% and 54% of

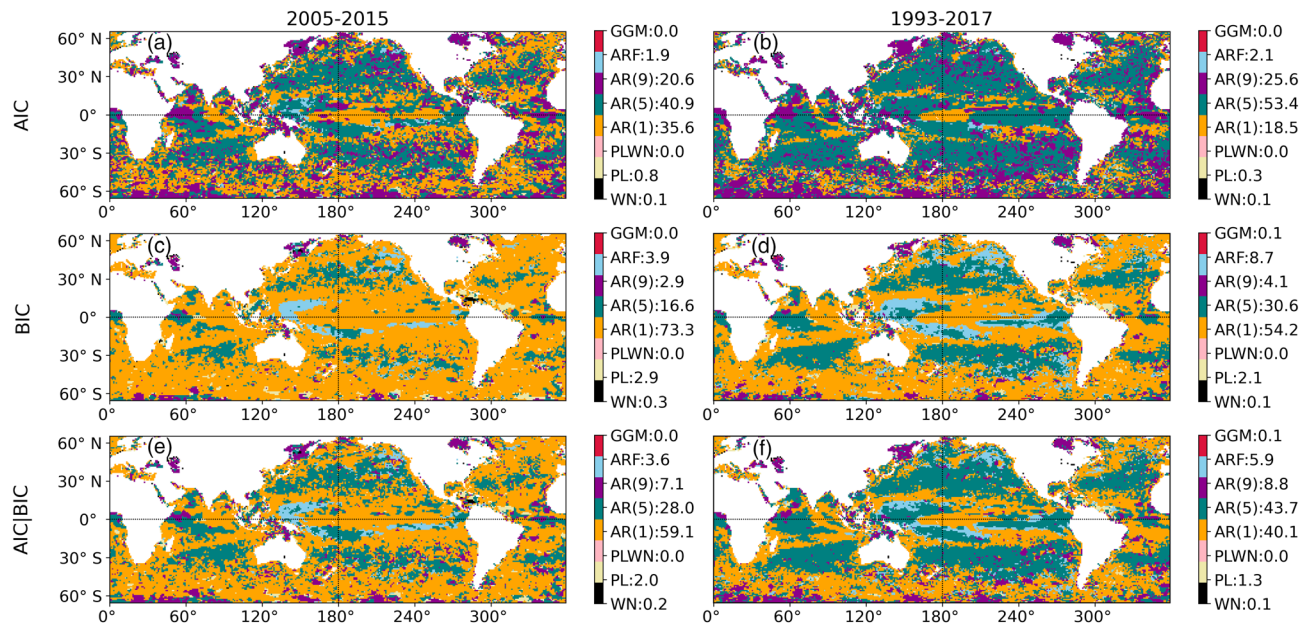


Figure 7. Preferred noise model for each grid for the ensemble mean with the respective percentage of the total ocean area selected by each noise model, based on (a, b) AIC, (c, d) BIC, and (e, f) the mean of AIC and BIC.

the ocean area for 2005–2015 and 1993–2017, respectively. For the 1993–2017 period, the AR(5) models is preferred in 30% of the ocean area, also according to BIC.

The noise model preference shows some distinctive regional patterns. The AR(5) pattern resembles the ocean gyres, suggesting a link between these dynamic regions and a preference for more complex noise models. The subtropical regions (both north and south) tend to also prefer the AR(5) model, for instance, in the AIC for 2005–2015 and BIC for 1993–2017. All scores show a clear “boomerang”-shaped pattern in the west equatorial Pacific (from $\sim 130^\circ\text{E}$ to 180°E of longitude and 20°S to 20°N of latitude), associated with the El Niño–Southern Oscillation (ENSO; Wang & Picaut, 2004). However, in this region different orders of autoregressive models score higher for the different criteria: AR(5) for AIC and ARF for BIC. While for some data sets, the GGM is selected as the preferred noise model for the global mean time series; regionally, this is almost never the case. In contrast, while the WN model is never preferred for the global mean, it is selected a few times in the regional patterns. Comparing the two time periods, we see an increase of the preference of higher order of autoregressive noise models with the increasing study period length, as already noted for the case of the global mean.

4.4. Regional Steric Sea-Level Trends

In the previous section, the regional noise model selection (Figure 7) showed that a combination of different noise models is necessary to best describe the regional ensemble mean trend and uncertainties. While mixing noise models is not standard practice, it allows local processes to be well represented in the trend and uncertainty estimations. Thus, in this section, we show the regional steric sea-level trend and uncertainty (Figure 8) using the preferred noise model for each grid point. As AR(1) was the preferred noise model in more than 50% of the ocean grids, the differences between the trends (Figures 8b and 8e) estimated with AR(1) and with the preferred noise model are only significant in the most dynamic regions. For both time periods, the differences become larger for the estimated uncertainties (Figures 8d and 8g).

Over the shorter time period (2005–2015), the regional trends are dominated by short-time-scale dynamics (Figures 8a and 8b). For example, the equatorial Pacific has a very strong signal of ENSO. The western boundary currents also display strong trend signals. A clear signal of the Gulf Current (west North Atlantic) is marked by a high trend of about 6 to 7 mm/year. The same is seen for the Kuroshio Extension (west North Pacific) and for the Malvinas Current (west South Atlantic). The Agulhas Current (South Africa) shows well-marked eddies. All these regions are dominated by internal variability rather than

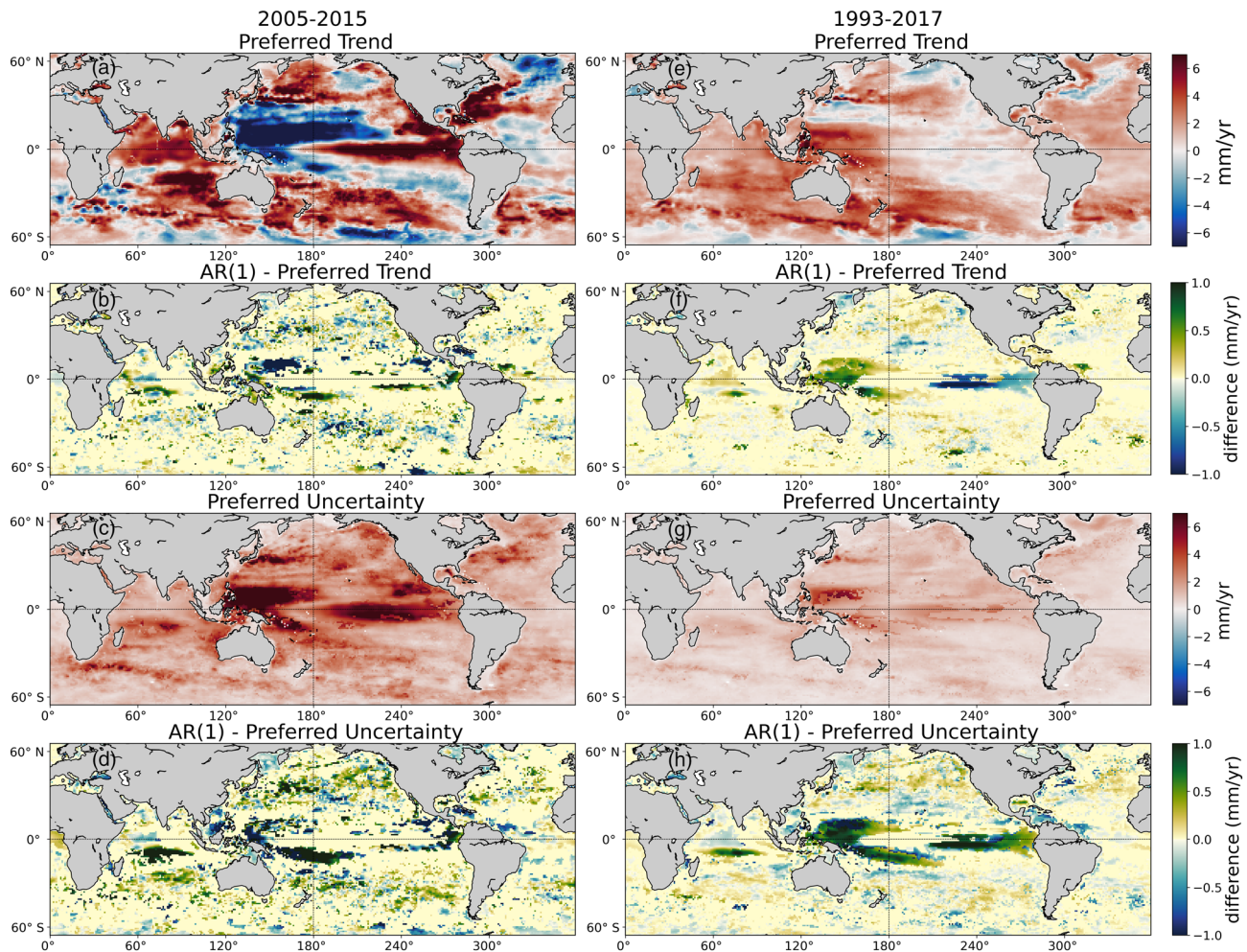


Figure 8. Trend and uncertainty (mm/year) for the ensemble mean using the preferred noise model for each grid (a, c, e, g) according to the mean of AIC and BIC (Figures 7e and 7f) and the differences between the estimates using AR(1) (Figure 5) and the preferred noise model (b, d, f, h), for (a–d) 2005–2015 and (e–g) 1993–2017.

a forced trend, with higher uncertainties (Figures 8c and 8d), and consequently, they are sometimes not statistically significant.

For the longer period (1993–2017), the trends (Figures 8e and 8f) are reduced in strength and less dominated by interannual oscillations and ocean currents, revealing a predominantly global positive trend. For instance, the strong positive trend in the east equatorial Pacific from ENSO fades into a weak negative trend, while the negative trend of the west is reversed to a positive, but weaker, trend. The trend also shows a weak reflection of the western boundary currents. The positive signal of the Gulf Stream is replaced by a negative trend. The negative trend of the North Atlantic Subpolar Gyre in 2005–2015 changes to a positive trend in 1993–2017, as also reported by Chafik et al. (2019). Although the Southern Ocean is outside our domain, we can still see in the lower boundaries of our maps the signal of the Antarctic Circumpolar Current, marked by a negative trend for both periods. As the time series get longer, internal variability has a smaller influence on the long-term trend, resulting in reduced uncertainties in 1993–2017 (Figures 8g and 8h) in comparison to those in 2005–2015.

5. Discussion

There is no clear consensus in the literature of which criterion should be used to select the best noise model, either AIC or BIC (Burnham and Anderson, 2002) or even another criterion (Klees & Broersen, 2002). While Sonnewald et al. (2018) use only the AIC, Hughes and Williams (2010) advise using BIC, which has less

spatial variability and thus would be more reliable than AIC. Figure 7 confirms this, since it displays more spatial differences for AIC selection than for BIC. However, if the aim is to investigate the spatial variability, then using a stricter criterion might not be preferable. Still, the preference of AIC for higher-order models and of BIC for simpler models is emphasized in Figures 3 and 7, as also reported by Sonnewald et al. (2018). To overcome these differences, we have chosen to compare and show both selection criteria, complemented by an ARMASA noise model analysis.

Our global mean analysis indicates that an AR(1) model is the best descriptor of the residuals of the ensemble mean steric SLC. This contrasts with the findings of Bos, Williams, et al. (2014), who investigated the effect of noise models on GMSL, focusing on altimetry and tide gauge sea-level data. They found that for altimetry data, ARF and GGM models are equally preferred, followed by an AR(5) model. Although preferable to a white noise model, the GGM model was one of the least preferred models in our global analysis. When considering all data sets in the noise model analysis, our results indicate that AR(5) and ARF are among the best candidates for the global mean time series, partially in line with the results of Bos, Williams, et al. (2014).

Regionally, our noise model analysis indicates that most of the oceans are best described by an AR(5) or AR(1) noise model. This agrees with the findings of Hughes and Williams (2010) and Royston et al. (2018), who both looked for the best noise model to describe regional sea level observed with tide gauges and satellite altimetry. While Hughes and Williams (2010) found that AR(5) is the best descriptor of the sea-level uncertainties, Royston et al. (2018) found that most of the regional monthly data (both from tide gauges and altimetry) can be described by the WN or AR(1) noise model. However, the WN model was almost never selected in our analysis. Both studies also highlight that the appropriateness of the noise model is ordered in larger spatial structures. The regional preference of the noise models (Figure 7) display spatial coherent patterns based on the temporal and spatial variability of the ocean processes. Our regional results agree to some extent with the spatial patterns described by previous studies, particularly the preference of more complex models in the tropics and highly dynamic regions. The latter might be related to multiple baroclinic modes present in these regions (Hughes & Williams, 2010).

The physics of the spatial patterns reported by Hughes and Williams (2010) were also confirmed by Sonnewald et al. (2018), who investigated the linear predictability of sea-surface height. Sonnewald et al. (2018) found that up to 50% of the sea-surface height variability over 20 years is explained by a seasonal signal. This might explain why the uncertainties for 1993–2017 are smaller than the ones for 2005–2015, as the linear trend becomes more important than the seasonal signal. The differences between noise model choices of previous studies (Hughes & Williams, 2010; Royston et al., 2018), which focused on tide gauge and altimetry data, and the present study, which focus on steric SLC, illustrate that the best noise model for steric SLC is not necessarily the best for total SLC. Hence, it is important to always analyze which noise model is most appropriate for each type of data and to consider the effect of different spatial (global vs. regional) and temporal (daily, monthly, or annual) resolutions.

Several sea-level budget studies use a suite of data sets, where the spread around the ensemble mean is used to describe the uncertainty of the sea-level trends (e.g., Cazenave et al., 2018; Gregory et al., 2013). This can result in misrepresentation of the uncertainty. For example, if we consider the 1-sigma spread of the estimated trends, then we would obtain an uncertainty of 0.71 and 0.36 mm/year, whereas with the AR(1) noise model we obtain uncertainties of 0.10 and 0.07 mm/year for 1993–2017 and 2005–2015, respectively. In some other cases, the authors do not mention how the uncertainties were obtained or what they represent in detail (e.g., Leuliette, 2015), making it difficult to interpret and compare steric estimates in the literature. As our results show, the noise model has an effect not only on the uncertainty but also on the derivation of the trend itself. This is expected once the stochastic model is incorporated in the least squares regression (Bos, Williams, et al., 2014). We caution that future studies should carefully describe how uncertainty estimates have been obtained, as this aids physical interpretation.

While we find that the noise model has the strongest effect on the uncertainty, our results show that differences in the trend are mainly a result of the data set choice, both at a regional scale (Figure S9) and for the global mean (Figure 2). Most of the differences was seen within the REA products, mainly the FOAM data set. A possible explanation can be the incorporation of temperature and salinity data from instrumented marine mammals (Carse et al., 2015), which may lead to a high bias in the salinity values. Furthermore,

the energy of deep currents (1,000 m) is highly overestimated in FOAM (Desportes et al., 2019). However, in comparison with other ocean REA, FOAM estimates are the closest to tropical mooring observations (Desportes et al., 2019) and should not be discarded.

The World Climate Research Programme (WCRP) sea-level budget report (Cazenave et al., 2018) studied the same Argo period, from January 2005 to December 2015, and found an ensemble mean thermosteric sea-level trend of 1.31 ± 0.40 mm/year based on 11 in situ data sets, which is considerably higher than our ensemble mean of 1.08 ± 0.07 mm/year. However, Cazenave et al. (2018) only included thermosteric SLC, and their rates are full depth: When excluding a deep ocean contribution of 0.1 ± 0.1 mm/year, their estimate is within 2-sigma of ours. Amin et al. (2020) considered five data sets (APDRC, SIORG, BOA, EN4, and JAMESTEC) to obtain a global mean steric sea-level trend for 2005–2016. Compared to our trends, both for individual data sets and for the ensemble, all the values agree within the uncertainty ranges, despite the extra year in their analysis.

For 1993–2015, the World Climate Research Programme report (Cazenave et al., 2018) found a thermosteric trend of 1.32 ± 0.4 mm/year, while Dieng et al. (2017) found a steric sea-level trend of 1.13 ± 0.12 mm/year for the same period, both based only on in situ data sets. Keeping in mind that these values are full depth and the period 2 years shorter than ours, they match our steric trend well for 1993–2017 (1.33 ± 0.11 mm/year). By merging several REA products, Storto et al. (2019) found a steric sea-level trend of 1.98 ± 0.3 mm/year for 1993–2016. Their trend agrees, within the uncertainty range, with our REA category mean, corroborating our findings. Although our trends agree within uncertainties with the values found in the literature, our uncertainty values are generally smaller than the previously reported ones, showing the added value of the noise model analysis.

Our results show that the variation of the steric sea-level trends from the use of different data sets can be reduced by using an ensemble of data sets. This is in line with the assumption that the systematic biases of the individual data sets are reduced by the averaging process (Storto et al., 2017). However, Rougier (2016) argues that the offsetting of biases alone is not a good reason to use the ensemble mean, as one should not expect that individual data sets will have such fundamental differences. Rougier (2016) poses that the ensemble mean does not contain all of the variability of the true process, which can result in underestimating the true sea-level trend and uncertainty. Instead of using the ensemble mean, one could use the data set with the lowest root-mean-square error in relation to the ensemble mean, which will retain the variability of an individual data set around the mean of the ensemble (Rougier, 2016; Royston et al., 2020). In our analysis, the ISAS+ and C-GLORS data sets are closest to the ensemble means for 2005–2015 and 1993–2017, respectively. For the global mean, the uncertainty of both data sets is best described by the AR(5) model. The regional pattern of preferred noise models (Figure S7) is dominated by AR(9) for 2005–2015 and by AR(5) for 1993–2017. Compared to the ensemble mean, the noise model selection according to ISAS+ and C-GLORS displays more spatial variability, especially for the shorter time period. By construction, the final trend and uncertainty patterns for ISAS+ and C-GLORS (Figure S8) are similar to the ensemble mean (Figure 8), though displaying higher spatial frequency features.

6. Conclusion

Motivated by the large discrepancies in dealing with uncertainties in the sea-level data sets, and the wide variety in steric sea-level trends published in the literature, this paper explored the variation in present-day steric SLC estimates. Two sources of variation were investigated: the uncertainties caused by the use of different data sets and by the use of different noise models to describe the residuals. We analyzed 15 data sets and eight noise models and showed the different rates and uncertainties of steric SLC for two time periods (2005–2015, the Argo period, and 1993–2017, the satellite altimetry era). By simultaneously varying all the data sets and noise models, the 2005–2015 global mean steric sea-level trend varied from 0.56 to 2.33 mm/year and the uncertainty from 0.02 to 1.65 mm/year.

Although the noise models are mainly used to describe the uncertainties of time series, we found that they also affect, to a smaller degree, the trend itself. By alternating the eight noise models for the ensemble mean, the 2005–2015 global mean steric sea-level trend varied from 1.07 ± 0.03 to 1.11 ± 0.19 mm/year. This illustrates the significant impact of the noise model on the uncertainty, which was even stronger on the regional scale. Our noise model analysis suggests that an AR(1) model is the most appropriate model to describe the

residual behavior of the global mean steric SLC of the ensemble mean and consequently leads to the most representative uncertainty. Regionally, a combination of different noise models is required to best describe steric SLC and its uncertainty. While AR(1) was the preferred noise model for most of the regions, ultimately the most appropriate noise model depends on the study location. The spatial coherence in the noise model preference shows clusters that have similar dynamics, which can be used to investigate the regional sea-level budget.

Compared to the noise model, the choice of the data set has a stronger influence on the estimated trend. By alternating the data sets while keeping the AR(1) noise model fixed, the steric SLC varied from 0.66 ± 0.12 to 1.86 ± 0.27 mm/year for 2005–2015. Using an ensemble mean of several data sets reduces the effect of uncertainties from a single data set and gives a more robust estimate of the observed steric change.

There are distinct differences in the trends between the three categories of data sets: REA, MiS, and only Argo. While the REA products have clear advantages, such as covering longer periods and depths than in situ-based data sets, there is a large spread of the results within the REA category. Using only one REA product to estimate the steric SLC might lead to a considerable overestimation or underestimation of the trend.

This study showed that the choice of data set and noise model results in large differences in the steric sea-level trend and the associated uncertainty. We therefore recommend that studies on sea-level trends and sea-level budgets perform a noise model analysis and report which noise model was used to determine the trend. An AR(1) noise model and an ensemble of data sets provide the best estimate of the steric contribution to global mean SLC up to at least 25 years. For the purpose of describing regional variations, more complex noise models are needed, such as higher-order autoregressive ones.

Data Availability Statement

The steric fields and trends computed in this research, as well as supporting scripts, are available at the 4TU. Research Data repository (<https://doi.org/10.4121/uuid:b6e5e4bc-d382-4b51-837b-c5cde4980bf3>). The temperature and salinity fields used are available at the hyperlinks of Table 1 and cited references.

Acknowledgments

This research was funded by the Netherlands Space Office User Support program (grant no. ALWO.2017.002). We thank Roland Klees for his input and support in the use of ARMASA and Soren Johnson for textual help. We used the software Hector (Bos et al., 2013) and ARMASA (Broersen, 2020) for the noise model analysis. The package TEOS-10 for Python was used to compute steric sea level. All figures in the manuscript were made in Python, with cmocean 2.0 colormaps (Thyng et al., 2016).

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Amante, C., & Eakins, B. W. (2009). ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. <https://doi.org/10.7289/V5C8276M> [24-04-2020].
- Amin, H., Bagherbandi, M., & Sjöberg, L. E. (2020). Quantifying barystatic sea-level change from satellite altimetry, GRACE and Argo observations over 2005–2016. *Advances in Space Research*, 65(8), 1922–1940. <https://doi.org/10.1016/j.asr.2020.01.029>
- Bindoff, N. L., Willebrand, J., Artale, V., Cazenave, A., Gregory, J. M., Gulev, S., et al. (2007). Observations: Oceanic climate change and sea level. In *Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 385–432). Cambridge: Cambridge University Press. Retrieved from <http://nora.nerc.ac.uk/id/eprint/15400>
- Blockley, E. W., Martin, M. J., McLaren, A. J., Ryan, A. G., Waters, J., Lea, D. J., et al. (2014). Recent development of the Met Office operational ocean forecasting system: An overview and assessment of the new Global FOAM forecasts. *Geoscientific Model Development*, 7(6), 2613–2638. <https://doi.org/10.5194/gmd-7-2613-2014>
- Bos, M. S., Fernandes, R. M. S., Vethamony, P., & Mehra, P. (2014). Estimating absolute sea level variations by combining GNSS and tide gauge data. *Indian Journal of Geo-Marine Sciences*, 43(7), 1327–1331.
- Bos, M. S., Fernandes, R. M. S., Williams, S. D. P., & Bastos, L. (2013). Fast error analysis of continuous GNSS observations with missing data. *Journal of Geodesy*, 87(4), 351–360. <https://doi.org/10.1007/s00190-012-0605-0>
- Bos, M. S., Williams, S. D. P., Araújo, I. B., & Bastos, L. (2014). The effect of temporal correlated noise on the sea level rate and acceleration uncertainty. *Geophysical Journal International*, 196(3), 1423–1430. <https://doi.org/10.1093/gji/ggt481>
- Broersen, P. M. T. (2002). Automatic spectral analysis with time series models. *IEEE Transactions on Instrumentation and Measurement*, 51(2), 211–216. <https://doi.org/10.1109/19.997814>
- Broersen, P. M. T. (2020). ARMASA. MATLAB central file exchange. Retrieved from <https://www.mathworks.com/matlabcentral/fileexchange/1330-armasa>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (Second ed., Vol. 53). New York: Springer. <https://doi.org/10.1017/CBO9781107415324.004>
- Cabanes, C., Grouazel, A., Von Schuckmann, K., Hamon, M., Turpin, V., Coatanoan, C., et al. (2013). The CORA dataset: Validation and diagnostics of in-situ ocean temperature and salinity measurements. *Ocean Science*, 9(1), 1–18. <https://doi.org/10.5194/os-9-1-2013>
- Carse, F., Martin, M. J., Sellar, A., & Blockley, E. W. (2015). Impact of assimilating temperature and salinity measurements by animal-borne sensors on FOAM ocean model fields. *Quarterly Journal of the Royal Meteorological Society*, 141(693), 2934–2943. <https://doi.org/10.1002/qj.2613>
- Carton, J. A., Chepurin, G. A., & Chen, L. (2018). SODA3: A new ocean climate reanalysis. *Journal of Climate*, 31(17), 6967–6983. <https://doi.org/10.1175/JCLI-D-18-0149.1>

- Carvalho, K. S., & Wang, S. (2019). Characterizing the Indian Ocean sea level changes and potential coastal flooding impacts under global warming. *Journal of Hydrology*, 569, 373–386. <https://doi.org/10.1016/j.jhydrol.2018.11.072>
- Cazenave, A., Meyssignac, B., Ablain, M., Balmaseda, M., Bamber, J., Barletta, V., et al. (2018). Global sea-level budget 1993–present. *Earth System Science Data*, 10(3), 1551–1590. <https://doi.org/10.5194/essd-10-1551-2018>
- Chafik, L., Nilsen, J. E. Ø., Dangendorf, S., Reverdin, G., & Frederikse, T. (2019). North Atlantic Ocean circulation and decadal sea level change during the altimetry era. *Scientific Reports*, 9, 1041. <https://doi.org/10.1038/s41598-018-37603-6>
- Cheng, L., Abraham, J., Goni, G., Boyer, T., Wijffels, S., Cowley, R., et al. (2016). XBT science: Assessment of instrumental biases and errors. *Bulletin of the American Meteorological Society*, 97(6), 924–933. <https://doi.org/10.1175/BAMS-D-15-00031.1>
- Cheng, L., Trenberth, K. E., Fasullo, J., Boyer, T., Abraham, J., & Zhu, J. (2017). Improved estimates of ocean heat content from 1960 to 2015. *Science Advances*, 3, e1601545. <https://doi.org/10.1126/sciadv.1601545>
- Church, J. A., Clark, P. U., Cazenave, A., Gregory, J. M., Svetlana Jevrejeva, A. L., Merrifield, M. A., et al. (2013). Sea level change. In T. F. Stocker, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1137–1216). Cambridge, U. K., New York: Cambridge Univ. Press.
- Desbruyères, D. G., Purkey, S. G., McDonagh, E. L., Johnson, G. C., & King, B. A. (2016). Deep and abyssal ocean warming from 35 years of repeat hydrography. *Geophysical Research Letters*, 43, 10,356–10,365. <https://doi.org/10.1002/2016GL070413>
- Desportes, C. C., Garric, G., Régnier, C., Drévilion, M., Parent, L., Garric, G., et al. (2019). For Global Ocean Reanalysis Multi-model Ensemble Products GREP.
- Dieng, H. B., Cazenave, A., Meyssignac, B., & Ablain, M. (2017). New estimate of the current rate of sea level rise from a sea level budget approach. *Geophysical Research Letters*, 44, 3744–3751. <https://doi.org/10.1002/2017GL073308>
- Dieng, H. B., Palanisamy, H., Cazenave, A., Meyssignac, B., & von Schuckmann, K. (2015). The sea level budget since 2003: Inference on the deep ocean heat content. *Surveys in Geophysics*, 36(2), 209–229. <https://doi.org/10.1007/s10712-015-9314-6>
- Frankcombe, L. M., Spence, P., Hogg, A. M., England, M. H., & Griffies, S. M. (2013). Sea level changes forced by Southern Ocean winds. *Geophysical Research Letters*, 40, 5710–5715. <https://doi.org/10.1002/2013GL058104>
- Gaillard, F., Reynaud, T., Thierry, V., Kolodziejczyk, N., & Von Schuckmann, K. (2016). In situ-based reanalysis of the global ocean temperature and salinity with ISAS: Variability of the heat content and steric height. *Journal of Climate*, 29(4), 1305–1323. <https://doi.org/10.1175/JCLI-D-15-0028.1>
- Garric, G., & Parent, L. (2017). Product user manual for global ocean reanalysis products GLOBAL-REANALYSIS-PHY-001-025. Retrieved from <http://cmems-resources.cls.fr/documents/PUM/CMEMS-GLO-PUM-001-025-011-017.pdf>
- Garry, F. K., McDonagh, E. L., Blaker, A. T., Roberts, C. D., Desbruyères, D. G., Frajka-Williams, E., & King, B. A. (2019). Model-derived uncertainties in deep ocean temperature trends between 1990 and 2010. *Journal of Geophysical Research: Oceans*, 124, 1155–1169. <https://doi.org/10.1029/2018JC014225>
- Gill, A. E., & Niller, P. P. (1973). The theory of seasonal variability in the ocean. *Deep-Sea Research*, 20, 141–177. [https://doi.org/10.1016/S0002-9610\(88\)80531-2](https://doi.org/10.1016/S0002-9610(88)80531-2)
- Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118, 6704–6716. <https://doi.org/10.1002/2013JC009067>
- Gouretski, V., & Reseghetti, F. (2010). On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database. *Deep-Sea Research Part I: Oceanographic Research Papers*, 57(6), 812–833. <https://doi.org/10.1016/j.dsr.2010.03.011>
- Gregory, J. M., White, N. J., Church, J. A., Bierkens, M. F. P., Box, J. E., van den Broeke, M. R., et al. (2013). Twentieth-century global-mean sea level rise: Is the whole greater than the sum of the parts? *Journal of Climate*, 26(13), 4476–4499. <https://doi.org/10.1175/JCLI-D-12-00319.1>
- Guinehut, S., Dhomp, A. L., Larnicol, G., & Le Traon, P. Y. (2012). High resolution 3-D temperature and salinity fields derived from in situ and satellite observations. *Ocean Science*, 8(5), 845–857. <https://doi.org/10.5194/os-8-845-2012>
- Hughes, C. W., & Williams, S. D. P. (2010). The color of sea level: Importance of spatial variations in spectral shape for assessing the significance of trends. *Journal of Geophysical Research*, 115, C10048. <https://doi.org/10.1029/2010JC006102>
- Ishii, M., & Kimoto, M. (2009). Reevaluation of historical ocean heat content variations with time-varying XBT and MBT depth bias corrections. *Journal of Oceanography*, 65(3), 287–299. <https://doi.org/10.1007/s10872-009-0027-7>
- Ishii, M., Yasui, S., Fukuda, Y., Sato, K., Hirahara, S., & Suzuki, T. (2017). Accuracy of global upper ocean heat content estimation expected from present observational data sets. *Solaia*, 13, 163–167. <https://doi.org/10.2151/sola.2017-030>
- Klees, R., & Broersen, P. (2002). *How to handle colored noise in large least-squares problems: Building the optimal filter*. Delft, The Netherlands: Delft University Press (DUP) Science.
- Kopp, R. E., Kemp, A. C., Bittermann, K., Horton, B. P., Donnelly, J. P., Cehrels, W. R., et al. (2016). Temperature-driven global sea-level variability in the Common Era. *Proceedings of the National Academy of Sciences*, 113(11), E1434–E1441. <https://doi.org/10.1073/pnas.1517056113>
- Leuliette, E. W. (2015). The balancing of the sea-level budget. *Current Climate Change Reports*, 1(3), 185–191. <https://doi.org/10.1007/s40641-015-0012-8>
- Li, H., Xu, F., Zhou, W., Wang, D., Wright, J. S., Liu, Z., & Lin, Y. (2017). Development of global gridded Argo data set with Barnes successive corrections. *Journal of Geophysical Research: Oceans*, 122, 866–889. <https://doi.org/10.1002/2016JC012264> Received
- Liddle, A. R. (2004). How many cosmological parameters? *Monthly Notices of the Royal Astronomical Society*, 351(3), L49–L53. <https://doi.org/10.1111/j.1365-2966.2004.08033.x>
- Llovel, W., Willis, J. K., Landerer, F. W., & Fukumori, I. (2014). Deep-ocean contribution to sea level and energy budget not detectable over the past decade. *Nature Climate Change*, 4(11), 1031–1035. <https://doi.org/10.1038/nclimate2387>
- Lu, S., Liu, Z., Li, H., Li, Z., Wu, X., Sun, C., & Xu, J. (2019). Manual of Global Ocean Argo gridded data set (BOA-Argo) (version 2019), 14 pp. <https://doi.org/10.1002/2016JC012285.6>
- MacIntosh, C. R., Merchant, C. J., & von Schuckmann, K. (2017). Uncertainties in steric sea level change estimation during the satellite altimeter era: Concepts and practices. *Surveys in Geophysics*, 38(1), 59–87. <https://doi.org/10.1007/s10712-016-9387-x>
- MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Ferday, D., Scaife, A. A., et al. (2015). Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141(689), 1072–1084. <https://doi.org/10.1002/qj.2396>
- McDougall, T. J., & Barker, P. M. (2011). Getting started with TEO-10 and the Gibbs Seawater Oceanographic Toolbox. Retrieved from http://www.teos-10.org/pubs/Getting_Started.pdf

- Purkey, S. G., & Johnson, G. C. (2010). Warming of global abyssal and deep Southern Ocean waters between the 1990s and 2000s: Contributions to global heat and sea level rise budgets. *Journal of Climate*, 23(23), 6336–6351. <https://doi.org/10.1175/2010JCLI3682.1>
- Roemmich, D., Johnson, G. C., Riser, S., Davis, R., Gilson, J., Owens, W. B., et al. (2009). The argo program: Observing the global ocean with profiling floats. *Oceanography*, 22(2), 34–43. <https://doi.org/10.5670/oceanog.2009.36>
- Roemmich, D., & Gilson, J. (2009). The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. *Progress in Oceanography*, 82(2), 81–100. <https://doi.org/10.1016/j.pocean.2009.03.004>
- Rougier, J. (2016). Ensemble averaging and mean squared error. *Journal of Climate*, 29(24), 8865–8870. <https://doi.org/10.1175/JCLI-D-16-0012.1>
- Royston, S., Vishwakarma, B. D., Westaway, R., Rougier, J., & Sha, Z. (2020). Can we resolve the basin-scale sea level trend budget from GRACE ocean mass? *Journal of Geophysical Research: Oceans*, 125, e2019JC015535. <https://doi.org/10.1029/2019JC015535>
- Royston, S., Watson, C. S., Legresy, B., King, M. A., Church, J. A., & Bos, M. S. (2018). Sea-level trend uncertainty with Pacific climatic variability and temporally-correlated noise. *Journal of Geophysical Research: Oceans*, 123, 1978–1993. <https://doi.org/10.1002/2017JC013655>
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. Retrieved from Accessed June 19, 2020. www.jstor.org/stable/2958889, <https://doi.org/10.1214/aos/1176344136>
- Slangen, A. B. A., Church, J. A., Zhang, X., & Monselesan, D. P. (2015). The sea level response to external forcings in historical simulations of CMIP5 climate models. *Journal of Climate*, 28(21), 8521–8539. <https://doi.org/10.1175/JCLI-D-15-0376.1>
- Sonnewald, M., Wunsch, C., & Heimbach, P. (2018). Linear predictability: A sea surface height case study. *Journal of Climate*, 31(7), 2599–2611. <https://doi.org/10.1175/JCLI-D-17-0142.1>
- Stammer, D., Cazenave, A., Ponte, R. M., & Tamisiea, M. E. (2013). Causes for contemporary regional sea level changes. *Annual Review of Marine Science*, 5(1), 21–46. <https://doi.org/10.1146/annurev-marine-121211-172406>
- Storto, A., Bonaduce, A., Feng, X., & Yang, C. (2019). Steric sea level changes from ocean reanalyses at global and regional scales. *Water*, 11, 1987. <https://doi.org/10.3390/w11101987>
- Storto, A., & Masina, S. (2016). C-GLORSv5: An improved multipurpose global ocean eddy-permitting physical reanalysis. *Earth System Science Data*, 8(2), 679–696. <https://doi.org/10.5194/essd-8-679-2016>
- Storto, A., Masina, S., Balmaseda, M., Guinehut, S., Xue, Y., Szekely, T., et al. (2017). Steric sea level variability (1993–2010) in an ensemble of ocean reanalyses and objective analyses. *Climate Dynamics*, 49(3), 709–729. <https://doi.org/10.1007/s00382-015-2554-9>
- Szekely, T., Gourrion, J., Pouliquen, S., & Reverdin, G. (2019). The CORA 5.2 dataset: Global in-situ temperature and salinity measurements dataset. Data description and validation. *Ocean Science Discussions*, 15, 1601–1614. <https://doi.org/10.5194/os-2018-144>
- Thyng, K. M., Greene, C. A., Hetland, R. D., Zimmerle, H. M., & DiMarco, S. F. (2016). True colors of oceanography: Guidelines for effective and accurate colormap selection. *Oceanography*, 29(3), 9–13. <https://doi.org/10.5670/oceanog.2016.66>
- Tomczak, M., & Godfrey, J. S. (2003). Temperature, salinity, density and the oceanic pressure field. *Regional oceanography: An introduction* (Chap. 2, 2nd Ed., pp. 15–28). Delhi: Daya Publishing House. Retrieved from <https://www.mt-oceanography.info/regoc/pdfversion.html>
- von Schuckmann, K., Palmer, M. D., Trenberth, K. E., Cazenave, A., Chambers, D., Champollion, N., et al. (2016). An imperative to monitor Earth's energy imbalance. *Nature Climate Change*, 6(2), 138–144. <https://doi.org/10.1038/nclimate2876>
- Wang, C., & Picaut, J. (2004). Understanding ENSO physics—A review. In C. Wang, S. Xie, J. Carton (Eds.), *Earth's climate: The ocean-atmosphere interaction, Geophysical Monograph Series* (Vol. 147, pp. 21–48). Washington, DC: American Geophysical Union. <https://doi.org/10.1029/147GM02>
- Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The ECMWF operational ensemble reanalysis-analysis system for ocean and sea ice: A description of the system and assessment. *Ocean Science*, 15(3), 779–808. <https://doi.org/10.5194/os-15-779-2019>

References From the Supporting Information

- Broersen, P. M. T. (2000). Finite sample criteria for autoregressive model order selection. *IEEE Transactions on Signal Processing*, 48(12), 3550–3558. <https://doi.org/10.1109/78.887047>