Biogeosciences

Open Access

EGU

Supplement of

Trend analysis of the airborne fraction and sink rate of anthropogenically released CO_2

Mikkel Bennedsen et al.

Correspondence to: Mikkel Bennedsen (mbennedsen@econ.au.dk)

1 Introduction

This document accompanies the paper "Trend analysis of the airborne fraction and sink rate of anthropogenically released CO_2" (Bennedsen et al., 2019): We expand on the analyses of the main paper and supply additional results.
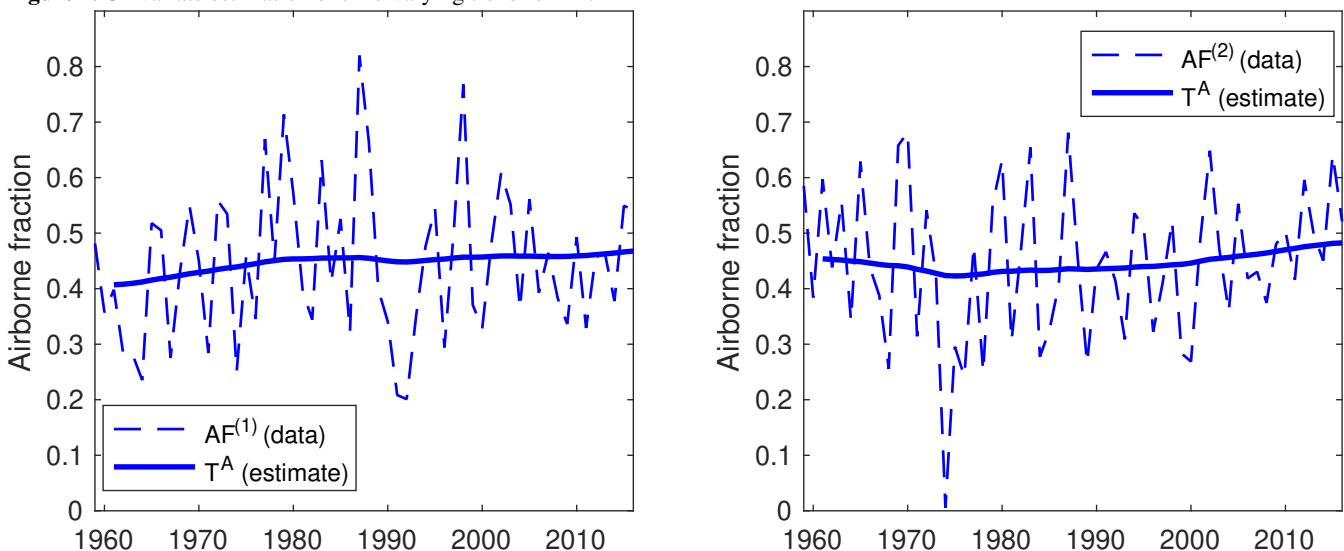
Section 2 provides some additional information of the univariate analyses of the main paper. In particular, Section 2.1 contains an analysis of potential statistically influential observations in the data set. Section 3 analyses data that have been averaged, so as to minimise possible influences from transitory events such as ENSO and volcanic eruptions.

2 Additional results from univariate analyses

In the main paper, we analyse the univariate data series in Tables 1 and 3 (airborne fraction and sink rate, respectively). However, to save space, we did not provide any graphical information on the estimated trends. These are given here, in Figure 1 (airborne fraction) and Figure 2 (sink rate). The figures can be compared to the ones from the multivariate analyses, see Figures 1 and 2 in the main paper. The main take-away from these figures is that in the univariate analyses, the underlying trends are found to be time-varying. This is contrast to the multivariate analyses, where the trends are better described by a trend that does not vary in time.

Figure 3 plots the trend estimates of the individual sink rates, i.e., of of the ocean sink rate (k_O) and land sink rate (k_L), cf. Equation (5) in the main paper.

Figure 1. Univariate estimation of time-varying trend for AF.

2.1 Influential data point analysis

To investigate the robustness of the statistical findings of the main paper, we here examine whether there are observations that are particularly influential on the analyses. We consider two related approaches. First, we calculate Cook's distance (Cook, 1977, 1979; Atkinson et al., 1997) for each $t = 1959, 1960, \ldots, 2016$, which is a measure of how influential a particular data point is. Second, we estimate the slope parameter β after deleting the t'th observation from the sample (treating it as a "missing value" in the state-space system) for $t = 1959, 1960, \ldots, 2016$. That is, for each year in the sample, we obtain an estimate $\beta_{\backslash t}$

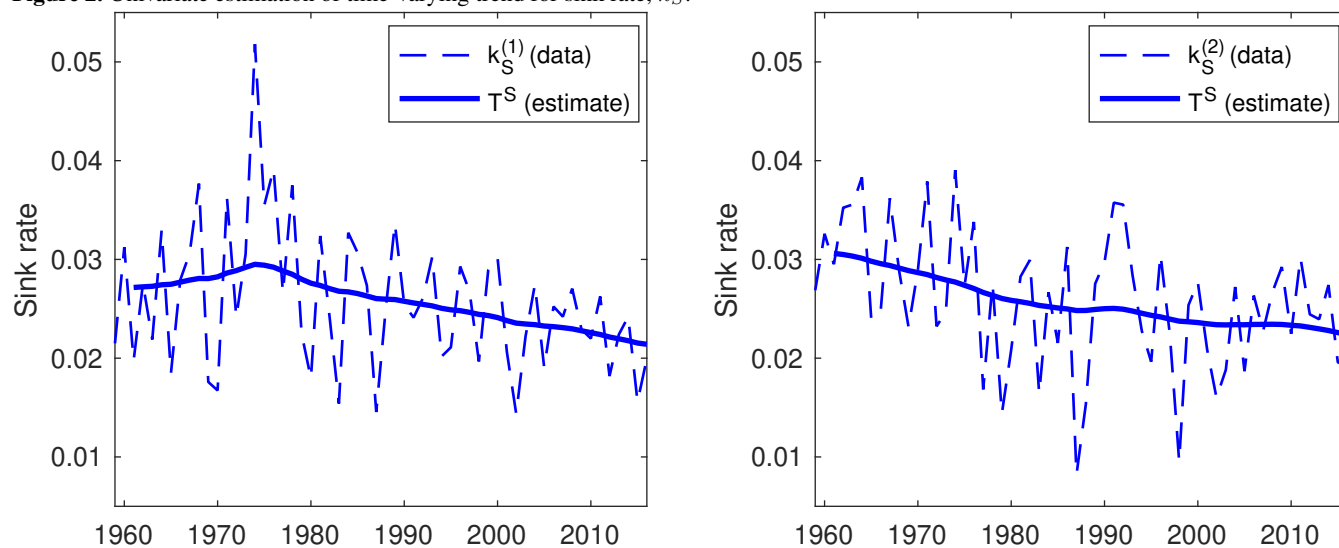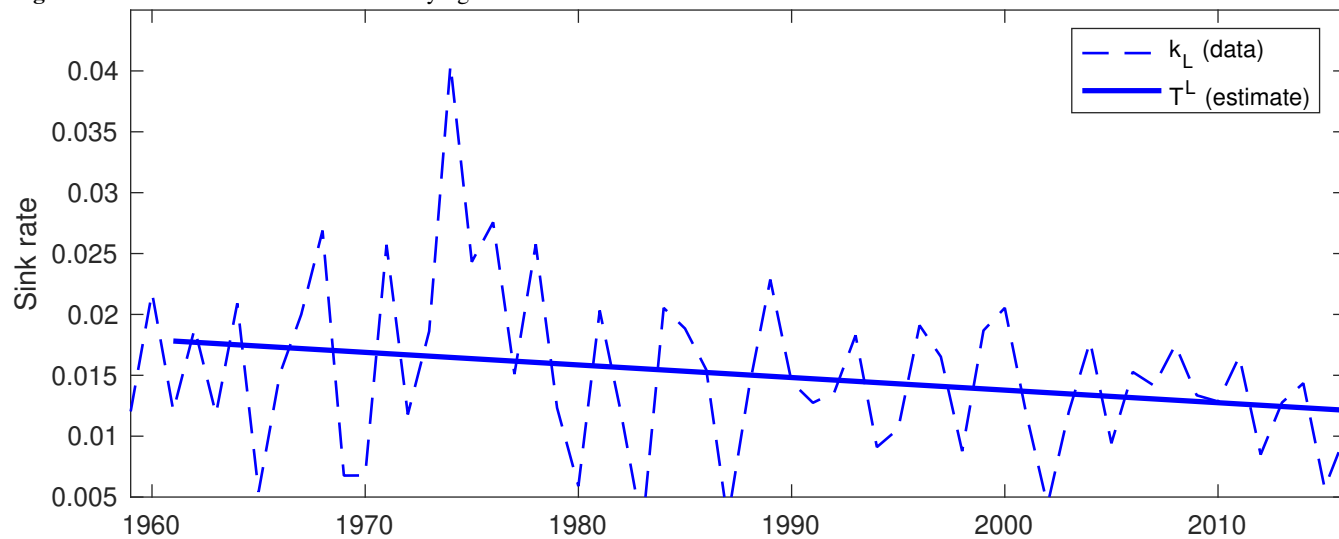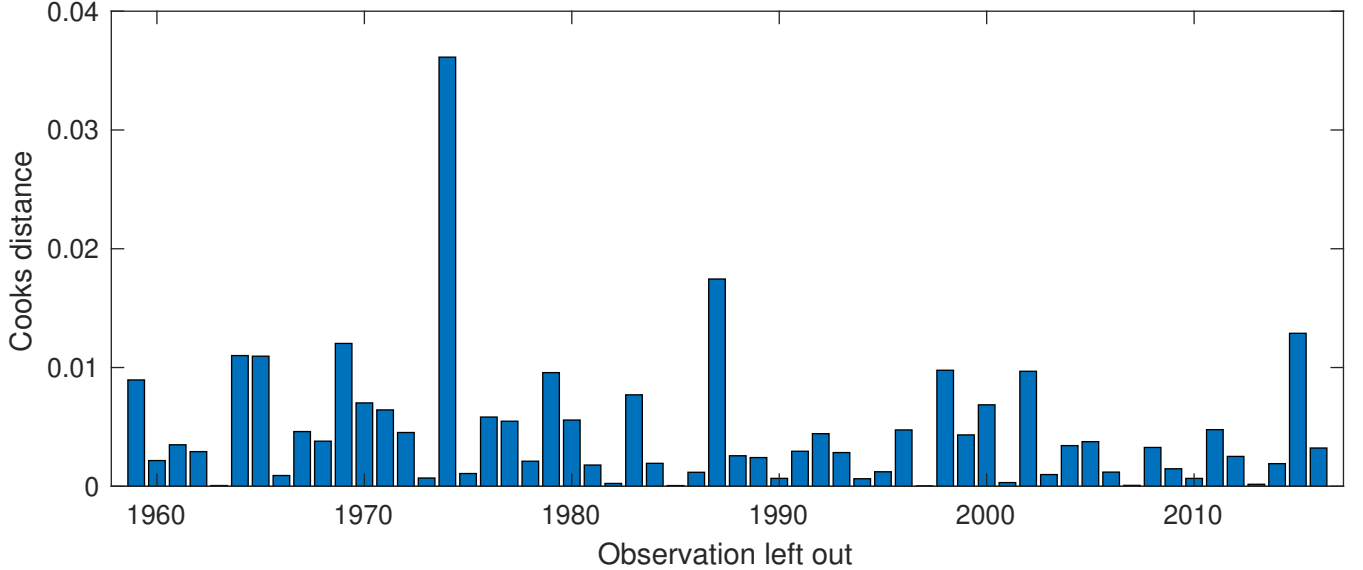**Figure 2.** Univariate estimation of time-varying trend for sink rate, k_S.

Figure 3. Univariate estimation of time-varying trend for land sink.

for each $t = 1959, 1960, \ldots, 2016$, by treating the t'th observation as missing and then estimating the slope parameter using the maximum likelihood approach considered in the main paper.

Figures 4 and 5 present these analyses for the airborne fraction, while Figures 6 and 7 present the analogous results for the sink rate. From these figures, it is clear that one observation, namely the one at $t = 1974$, stands out: it has by far the largest
5   Cook's distance for both the airborne fraction and the sink rate data (Figures 4 and 6, respectively). It also results in the most extreme slope parameter estimates (Figures 5 and 7, respectively). However, even this observation at $t = 1974$ seems to have only a minor influence on the analysis: indeed, the estimate of $\beta_{\backslash 1974}$ is well within the confidence bands (dashed lines) of the full sample estimate for both the airborne fraction data (Figure 5) and for the sink rate data (Figure 7). Similarly, while the Cook's distance for this data point is high relative to the other points, it is far below conventional rule-of-thumb thresholds
10   such as 0.50 or 1.00, which are often used in applied analyses.

Figure 4. Cook's distance of each observation from the state-space model of the airborne fraction.

3   Analysis with 5-year averages

To remove possible influences of transitory events (e.g., ENSO and/or volcanic eruptions) on our analyses, we here re-do the analysis from the main paper, now with data averaged over 5 year periods.[1] We also averaged the data over 2, 3, and 4 years and found similar results to what we report below. For brevity, the results from using 2-4 year averages are not presented. They
15   are available upon request.

We consider two ways of averaging the data: (i) Running averages, which results in overlapping "windows" of data, and (ii) non-overlapping averages. The former is considered in Section 3.1 and the latter in Section 3.2.

3.1   Running averages

Recall that the data of the original paper runs from $t = 1959$ to $t = 2016$, resulting in 58 observations. We now study the data
20   after they have been averaged over a 5-year period using a running window. That is, if the original data are x_t, we consider the

---

[1] We thank an anonymous referee for suggesting this.

Figure 5. Estimates of slope parameter β from the state-space model of the airborne fraction after leaving out one observation. Dashed lines are 95% confidence bands for the full sample estimate.
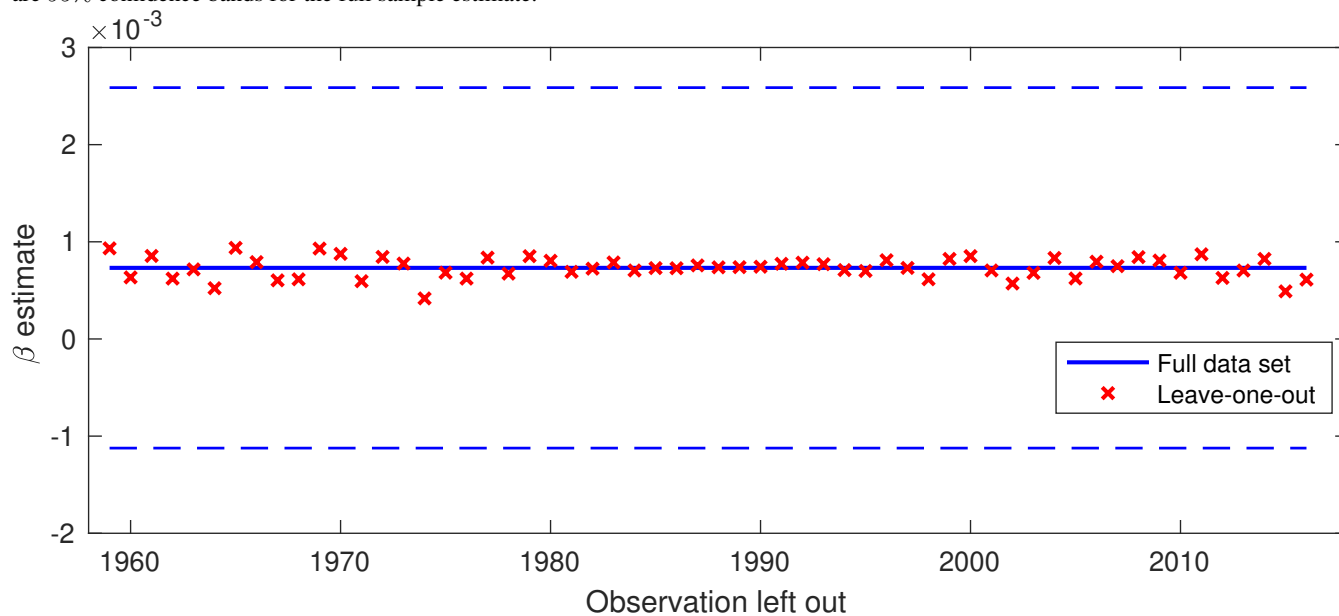
Figure 6. Cook's distance of each observation from the state-space model of the sink rate.
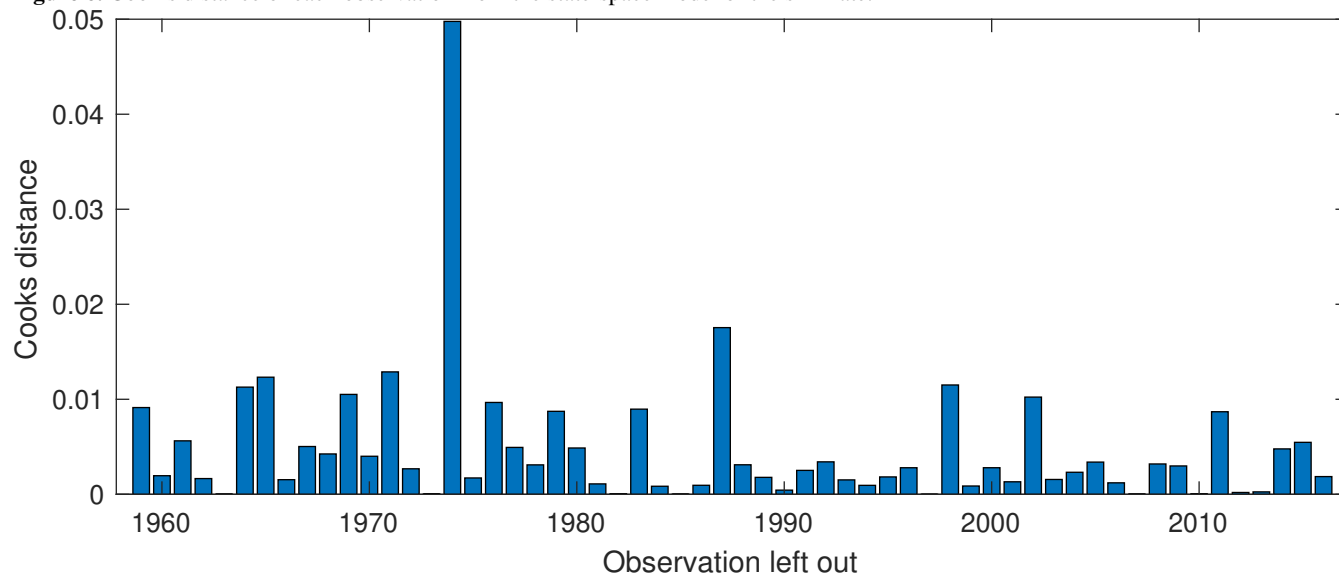
Figure 7. Estimates of slope parameter β from the state-space model of the sink rate after leaving out one observation. Dashed lines are 95% confidence bands for the full sample estimate.

averaged data

$$\tilde{x}_t^{(m)} = \frac{1}{m}\sum_{i=1}^{m} x_{t-i+1}, \quad t = 1958+m, 1959+m, \ldots, 2016,$$

where $m \geq 1$ is the number of observations used in constructing the averages. For instance, if $m = 5$, we get the 5-year averages

$$\tilde{x}_t^{(5)} = \frac{1}{5}\sum_{i=1}^{5} x_{t-i+1}, \quad t = 1963, 1963, \ldots, 2016,$$

5 resulting in $58 - 4 = 54$ observations.

As remarked in the Discussion section of the main paper, such averaging can, unfortunately, make the error structure of the data quite complicated. We illustrate how this can happen with a toy example: Suppose that the original data are "signal plus noise":

$$x_t = x_t^* + \xi_t,$$

10 where x_t^* is the true (unobserved) value for the underlying data and ξ_t is an iid measurement error term. Let $m \geq 2$ and consider the averaged data

$$\tilde{x}_t^{(m)} = \frac{1}{m}\sum_{i=1}^{m} x_{t-i+1} = \tilde{x}_t^* + \tilde{\xi}_t,$$

where

$$\tilde{x}_t^* = \frac{1}{m}\sum_{i=1}^{m} x_{t-i+1}^*, \qquad \tilde{\xi}_t = \frac{1}{m}\sum_{i=1}^{m} \xi_{t-i+1},$$

are the averaged signal and an error term, respectively. It is clear that the error term $\tilde{\xi}_t$ is now serially correlated, which can invalidate the analysis if the researcher does not take it into account.

With the above caveat in place, we analyse the 5-year averaged airborne fraction (AF) and sink rate (SR) data using a trend model specification, as explained in Section 3 of the main paper.[2] The results are shown in Tables 1 and 2 for the airborne
5 fraction and Tables 3 and 4 for the sink rate. From the tables, we see that the diagnostics are quite bad, indicating that the proposed model is not able to fit the data well. In particular, we find evidence of positive serial correlation ($DW < 2$) in the prediction errors, as we would expect from the discussion of the error structure above.

We conclude that if one wants to analyse the running averages data, another approach than the one considered here is necessary. In particular, it is important to choose an approach that can take into account the error structure induced by the
10 averaging. However, constructing such an alternative approach is outside the scope of this note. Instead, we propose to average the data using non-overlapping windows, thereby hopefully alleviating the serial correlation in the errors. The following section presents the results from this approach.

Table 1. Univariate analysis of the airborne fraction

	Parameter estimates					Diagnostics		
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	N	R_d^2	DW
$AF_t^{(1)}$	0	0.0447	0.00223	0.00584	0.38270	0.3387	-0.0616	1.6657
$AF_t^{(2)}$	0	0.0444	0.00055	0.00610	0.08987	1.0360	-0.0711	1.8195

Data: 5-year running average AF. See Table 1 in the main paper for the analogous analysis on the original data.

Table 2. Multivariate analysis of the airborne fraction

	Parameter estimates					Correlation matrix (ϵ)		Diagnostics		
Panel A: Two individual trends as in Eq. (12) of the main paper.										
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	$AF^{(1)}$	$AF^{(2)}$	N	R_d^2	DW
$AF^{(1)}$	0	0.0425	0.00223	0.00584	0.38270	1.0000	0.9999	0.3387	-0.0616	1.6657
$AF^{(2)}$	0	0.0444	0.00055	0.00610	0.08987	0.9999	1.0000	1.0360	-0.0711	1.8195
Panel B: One common trend as in Eq. (13) of the main paper.										
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	$AF^{(1)}$	$AF^{(2)}$	N	R_d^2	DW
$AF^{(1)}$	0.0359	0.0393	0.00149	0.00540	0.27679	1.0000	-1.0000	3.6409	-0.6278	1.1458
$AF^{(2)}$	0.0459	–	–	–	–	-1.0000	1.0000	0.8089	-0.9646	0.9314

Data: 5-year running average AF. See Table 2 in the main paper for the analogous analysis on the original data.

15

---

[2]We perform the averaging directly on the global carbon budget data, i.e., on $E_t^{ANT}, G_t, S_t^O, S_t^L$, and C_t, and then construct the AF and SR data as explained in Section 2 of the main paper. We also experimented with first constructing the AF and SR data and then averaging these. The results of the two approaches were very similar so we only present the results of the former.

Table 3. Univariate analysis of the sink rate

	Parameter estimates					Diagnostics		
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	N	R_d^2	DW
$k_S^{(1)}$	0	0.0022	-0.00008	0.00030	-0.27290	4.3025	-0.0710	1.7625
$k_S^{(2)}$	0	0.0020	-0.00017	0.00027	-0.63234	0.0726	-0.0637	1.7127

Data: 5-year running average SR. See Table 3 in the main paper for the analogous analysis on the original data.

Table 4. Multivariate analysis of the sink rate

	Parameter estimates					Correlation matrix (ϵ)		Diagnostics		
Panel A: Two individual trends as in Eq. (14) of the main paper.										
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	$AF^{(1)}$	$AF^{(2)}$	N	R_d^2	DW
$k_S^{(1)}$	0	0.0022	-0.00008	0.00030	-0.27290	1.0000	1.0000	4.3025	-0.0710	1.7625
$k_S^{(2)}$	0	0.0020	-0.00017	0.00027	-0.63235	1.0000	1.0000	0.0726	-0.0637	1.7127
Panel B: One common trend as in Eq. (15) of the main paper.										
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	$AF^{(1)}$	$AF^{(2)}$	N	R_d^2	DW
$k_S^{(1)}$	0.0024	0.0019	-0.00014	0.00026	-0.53638	1.0000	-1.0000	0.6056	-0.9981	0.8756
$k_S^{(2)}$	0.0015	–	–	–	–	-1.0000	1.0000	5.6559	-0.6044	1.1971

Data: 5-year running average SR. See Table 4 in the main paper for the analogous analysis on the original data.

3.2 Non-overlapping averages

We here consider averaged data, but where the averaging is done using non-overlapping windows. That is, if x_t is again the original data series, we define

$$\tilde{x}_t^{(m)} = \frac{1}{m}\sum_{i=1}^{m} x_{t-i+1}, \quad t = 1958+m, 1958+2m, \ldots, 2016,$$

5 i.e., we divide the data into bins ("windows") of size m and then take the average in each bin.[3]

The advantage of this approach to averaging the data is, as discussed above, that we expect the errors to be more nicely behaved. In particular, we do not expect them to be as serially correlated in this case. The downside of the approach is of course that we will have fewer observations available for the subsequent statistical analysis. For instance, in the case of 5-year averages that we consider here, each averaged data series consists of only 12 observations.

10 The results are shown in Tables 5 and 6 for the airborne fraction and Tables 7 and 8 for the sink rate. From the tables, we see, as expected, that the diagnostics are better than what was seen in the case of overlapping averaging windows. Interestingly, the two main conclusions from the main paper hold also in the case of the averaged data: (i) We find no statistical evidence of an increasing airborne fraction (Tables 5 and 6) but we do find statistical evidence of a decreasing sink rate (Panel B of Table 6). (ii) The latter conclusion is only reached when *all* the data are combined into one model with a common trend, compare Panel

15 A and Panel B of Table 6.

Figures 8 and 9 show the trend estimates of the airborne fraction and sink rate data, respectively, in the case of averaged data using non-overlapping windows.

---

[3]Because the original data set consists of 58 observations, it might not be possible in practice to have exactly m observations in each bin; in this case the first bin will have less observations allocated to it. For instance, in the case of 5-year averaging, the first bin will consist of 3 observations (representing the years 1959, 1960, and 1961) and the other bins will have 5 observations allocated to them (i.e., the second bin consists of the years 1962-1966, the third bin 1967-1971, etc.).

Table 5. Univariate analysis of the airborne fraction

	Parameter estimates					Diagnostics		
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	N	R_d^2	DW
$AF_t^{(1)}$	0.0574	0.0067	0.00517	0.00523	0.98975	0.3594	0.3465	2.0985
$AF_t^{(2)}$	0.0481	0.0450	0.00083	0.01449	0.05721	0.1227	0.1018	2.0714

Data: 5-year non-overlapping average AF. See Table 1 in the main paper for the analogous analysis on the original data.

Table 6. Multivariate analysis of the airborne fraction

	Parameter estimates					Correlation matrix (ϵ)		Diagnostics		
Panel A: Two individual trends as in Eq. (12) of the main paper.										
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	$AF^{(1)}$	$AF^{(2)}$	N	R_d^2	DW
$AF^{(1)}$	0.0473	0.0335	0.00538	0.01087	0.49443	1.0000	0.9456	0.7162	0.4872	2.5460
$AF^{(2)}$	0.0394	0.0541	0.00122	0.01666	0.07344	0.9456	1.0000	0.3027	0.0821	2.3698
Panel B: One common trend as in Eq. (13) of the main paper.										
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	$AF^{(1)}$	$AF^{(2)}$	N	R_d^2	DW
$AF^{(1)}$	0.0535	0	0.00361	0.00364	0.99164	1.0000	0.1975	0.4802	0.3698	2.2944
$AF^{(2)}$	0.0596	–	–	–	–	0.1975	1.0000	0.7454	0.5965	2.6334

Data: 5-year non-overlapping average AF. See Table 2 in the main paper for the analogous analysis on the original data.
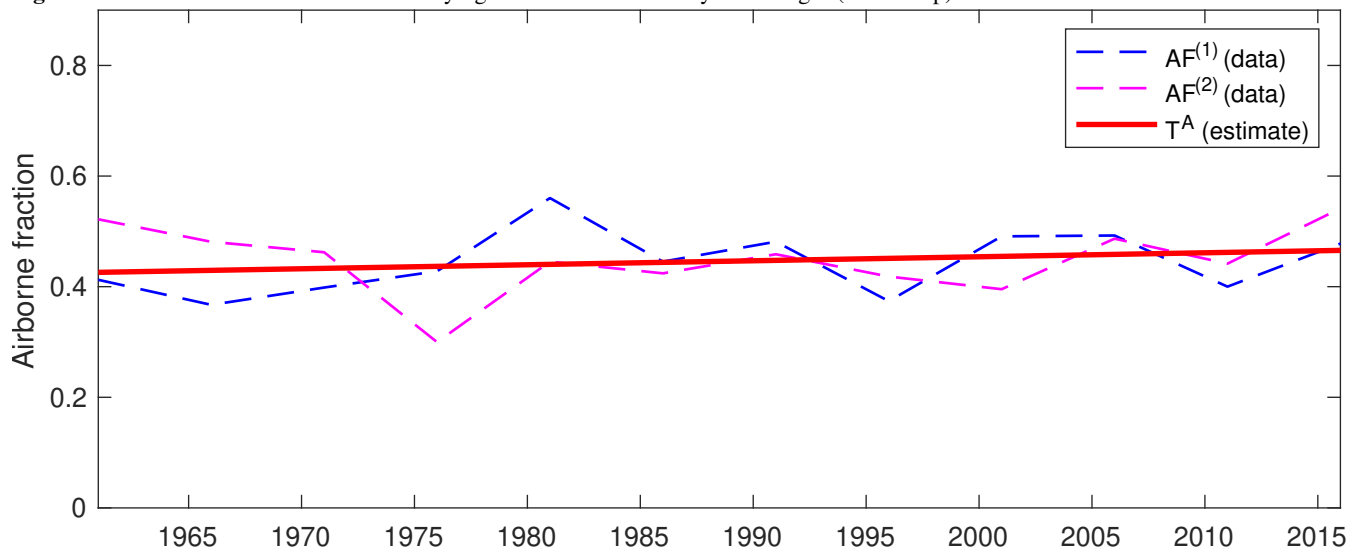
Figure 8. Multivariate estimation of time-varying trend for AF. Data: 5-year averages (no overlap)

Table 7. Univariate analysis of the sink rate

	Parameter estimates					Diagnostics		
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	N	R_d^2	DW
$k_S^{(1)}$	0.0023	0.0029	-0.00037	0.00091	-0.40483	0.1374	-0.0147	2.0028
$k_S^{(2)}$	0.0026	0.0011	-0.00073	0.00041	-1.78102	0.0770	0.2431	2.0450

Data: 5-year non-overlapping average SR. See Table 3 in the main paper for the analogous analysis on the original data.

Table 8. Multivariate analysis of the sink rate

	Parameter estimates					Correlation matrix (ϵ)		Diagnostics		
Panel A: Two individual trends as in Eq. (14) of the main paper.										
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	$AF^{(1)}$	$AF^{(2)}$	N	R_d^2	DW
$k_S^{(1)}$	0.0023	0.0029	-0.00052	0.00088	-0.58373	1.0000	1.0000	0.7418	-0.0204	2.0524
$k_S^{(2)}$	0.0025	0.0015	-0.00078	0.00049	-1.57493	1.0000	1.0000	0.0552	0.4228	2.2251
Panel B: One common trend as in Eq. (15) of the main paper.										
	$\widehat{\sigma}_\epsilon$	$\widehat{\sigma}_\eta$	$\widehat{\beta}$	s.e.$(\widehat{\beta})$	t-stat$(\widehat{\beta})$	$AF^{(1)}$	$AF^{(2)}$	N	R_d^2	DW
$k_S^{(1)}$	0.0032	0	-0.00068	0.00020	-3.41506	1.0000	0.3340	0.1292	0.6888	2.9392
$k_S^{(2)}$	0.0027	–	–	–	–	0.3340	1.0000	1.0816	0.1869	1.8314

Data: 5-year non-overlapping average SR. See Table 4 in the main paper for the analogous analysis on the original data.

References

Atkinson, A. C., Koopman, S. J., and Shephard, N.: Detecting shocks: Outliers and breaks in time series, Journal of Econometrics, 80, 387–422, 1997.

Bennedsen, M., Hillebrand, E., and Koopman, S. J.: Trend analysis of the airborne fraction and sink rate of anthropogenically released CO_2, Biogeosciences, 2019.

Cook, R. D.: Detection of influential observations in linear regression, Technometrics, 19, 15–18, 1977.

Cook, R. D.: Influential observations in linear regression, Journal of the American Statistical Association, 74, 169–174, 1979.

5

Figure 9. Multivariate estimation of time-varying trend for sink rate, k_S. Data: 5-year averages (no overlap)