Response to Referee Report

L. M. W. Leggett and D. A. Ball*
*Global Risk Policy Group Pty Ltd, Townsville, Queensland, Australia

Abstract

Response to Referee Report on "First and second derivative atmospheric CO_2, global surface temperature and ENSO"

----

1 Overall Response

We would like to thank the anonymous referee for providing comments on our submission. In overall response to the comments we note that we found all the points most useful and leading to the running of further checks or to important clarifications or additions. We now address the referee’s individual comments in turn.

2 Responses to Individual Comments

Comment 1

The issue that the series for temperature and CO_2 since 1850 exhibit different degrees of integration, and hence cannot modelled conventionally, was the subject of an important paper by Beenstock et al. (Earth System Dynamics 3 (2012), pp 173-188). These authors studied annual data, and concluded that the series over the 1850-2007 period were best described as integrated of order 1 (I(1)) in the case of temperature and I(2) in the case of CO_2*. They therefore conducted a cointegration analysis between temperature and ï ¿AD*CO_2 *(ï ¿Aˇ D denoting first*-*differencing), rather than a correlationˇ analysis, as appears here. Both studies therefore focus on dealing with the fact that a statistical model linking the levels of CO_2 and temperature cannot be constructed. However, differences of timespan, and data frequency, lead them to different interpretations of this fact, which is an issue that deserves careful consideration, in itself. It is clear, in any case, that the present authors must reference the Beenstock et al. study, and reconcile their findings with the previous reported ones.*

First, we agree that reference should be made to Beenstock et al.

Second, the essential point for the present study is that Beenstock et al. (2012) show in their work that the order of integration for temperature is I(1) while that for first-difference (equivalent to first-derivative) atmospheric CO_2 is I(2). In our paper we provide evidence that first-derivative atmospheric CO_2 is I(1).

Concerning the reconciliation of these two varying results, Pretis and Hendry (2013) have reviewed Beenstock et al. (2012). They take issue with the finding of I*(2)*, and find evidence that it results from the combination of two different data sets measured in different ways to make up the tested 1850-2011 data set which Beenstock et al. use. Concerning this composite series they write:

> In the presence of these different measurements exhibiting structural changes, a unit-root test on the entire sample could easily not reject the null hypothesis of (2) even when the data are in fact I*(1)*. Indeed, once we control for these changes, our results contradict the findings in Beenstock et al. (2012).

To focus on the first-derivative CO_2 data, which is relevant to our paper, we note that Pretis and Hendry (2013) show that, when the series are broken up into their two underlying series each measured in its own way and assessed using the ADF procedure (Response Table 1) the null hypothesis (that the first-derivative CO_2 series is non-stationary) is rejected.

Response Table 1: Table 1 from Pretis and Hendry (2013)

Table 1. ADF unit-root tests on $\Delta rfCO_2$.

1850–1957 constant			1958–2011 constant and trend		
D-lag	t-adf	Reject H_0	D-lag	t-adf	Reject H_0
5★★	−3.737	**	5	−4.089	*
4	−2.910	*	4	−3.807	*
3	−2.948	*	3	−3.383	
2	−3.146	*	2	−4.197	**
1	−2.706		1	−5.365	**
0	−3.544	**	0	−6.563	**

ADF unit-root tests: the null hypothesis H_0 is that the series has a unit root so is non-stationary. Rejecting the null hypothesis suggests no unit-root non-stationarity. D-lag specifies the number of lags included in the ADF unit root test, where ★ indicates that longest lag is significant at 5 % and ★★ at 1 %. If no lags are significant, the model with zero lags is appropriate. Unit root test outcome: ** indicates rejection of the null hypothesis at 1 % and * at 5 %.

Our results for CO_2 use instrumental data from the period 1958, matching one of the two time periods covered in Pretis and Hendry (2013) Table 1 above.

For this period in the paper we used monthly data. Here we provide that again (in Response Table 2) and also repeat the analysis for annual data (Response Table 3):

Response Table 2 - monthly data

Augmented Dickey-Fuller test for N2x13mma_1stderivCO_2 including 8 lags of (1-L)N2x13mma_1stderivCO_2

(max was 10, criterion modified AIC)
sample size 635
unit-root null hypothesis: a = 1

test without constant
model: (1-L)y = (a-1)*y(-1) + ... + e
1st-order autocorrelation coeff. for e: 0.005
lagged differences: F(8, 626) = 87.259 [0.0000]
estimated value of (a - 1): -0.0131027
test statistic: tau_nc(1) = -3.03873
asymptotic p-value 0.002319

test with constant
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
1st-order autocorrelation coeff. for e: 0.005
lagged differences: F(8, 625) = 87.229 [0.0000]
estimated value of (a - 1): -0.0143456
test statistic: tau_c(1) = -3.06294
asymptotic p-value 0.02944

with constant and trend
model: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
1st-order autocorrelation coeff. for e: 0.005
lagged differences: F(2, 636) = 292.044 [0.0000]
estimated value of (a - 1): -0.0319119
test statistic: tau_ct(1) = -5.02465
asymptotic p-value 0.0001

Response Table 3 - annual data

Augmented Dickey-Fuller test for Atmos_CO_2 including 6 lags of (1-L)Atmos_CO(max was 10, criterion modified AIC)

sample size 48
unit-root null hypothesis: a = 1

test without constant
model: (1-L)y = (a-1)*y(-1) + ... + e
1st-order autocorrelation coeff. for e: -0.035
lagged differences: F(6, 41) = 7.726 [0.0000]
estimated value of (a - 1): 0.0620622
test statistic: tau_nc(1) = 1.37673
asymptotic p-value 0.9583

test with constant
model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
1st-order autocorrelation coeff. for e: 0.008
lagged differences: F(9, 34) = 2.467 [0.0276]
estimated value of (a - 1): -0.164902
test statistic: tau_c(1) = -0.789087
asymptotic p-value 0.8217

with constant and trend
model: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
1st-order autocorrelation coeff. for e: -0.001
lagged differences: F(3, 45) = 0.695 [0.5601]
estimated value of (a - 1): -1.09988
test statistic: tau_ct(1) = -3.42433
asymptotic p-value 0.04814

Comparison of the relevant sections of Response Tables 1, 2 and 3 shows that (i) our results for annual data replicate those of Pretis and Hendry (2013) closely, and that (ii) the use of monthly data increases the statistical significance of the (already statistically significant) result substantially, by some two orders of magnitude.

It is hoped that the above reconciles our findings with those in the literature and shows that it is appropriate to treat first-difference CO_2 as I(1) for the period 1959 onward.

(For suggested changes to text from Comment 1 see response to Comment 2 next.)

Comment 2

In fact, there is a considerable degree of controversy (see for example the comments on the Beenstock paper in ESD) about the order of integration of these series, and as to whether they are stochastic trend processes (I(1) or I(2)) or "trend stationary" over sub-periods, with periodic breaks in trend. The essential problem here, I think, is that the time series models invoked in the literature on nonstationarity are rather *simple, and cannot play the role of what econometricians call the "data generation process". At best, they are simplified descriptions that apply only over limited spans of time. This fact throws conventional inference procedures (which have a large-sample justification) into some doubt.*

T he answer provided under Comment 1 addresses most of the points related to Beenstock et al. (2012), However we would also suggest adding the following.

> "The frequency of the data is unlikely to account for this difference in the results. This is because the (true) order of integration of a time-series is invariant to temporal aggregation; and the ability of the ADF test to detect this order is also unaffected by the sampling frequency, especially with relatively large sample sizes (*e.g.*, Pierce and Snell, 1995)."

Specifically addressing the comment "…..conventional inference procedures (which have a large-sample justification)…", it is noted that most of the inferential procedures we use are valid in finite samples, as well as asymptotically. For example, in the case of ADF testing, *exact* critical values are used.

Suggested changes to the paper

To deal with Comments 1 and 2 overall we suggest the addition of the following paragraph on page 29117, before the paragraph that starts on line 11 with the words "In contrast…":

> In carrying this out, one must first note that while we find, as is required for time series analysis, that the variables TEMP and FIRSTDERIVATIVE CO_2 are both stationary, (that is, both display order of integration of *I (1)*), Beenstock et al. (2012) report in their work that temperature *is I(1)* while first-difference (equivalent to first-derivative) atmospheric CO_2 is *I*(2).
>
> With regard to the reconciliation of these two varying results, Pretis and Hendry (2013) have reviewed Beenstock et al. (2012). They take issue with the finding of *I(2)*, and find evidence that it results from the combination of two different data sets measured in different ways to make up the tested 1850-2011 data set which Beenstock et al. use. Regarding this composite series they write:
>
>> In the presence of these different measurements exhibiting structural changes, a unit-root test on the entire sample could easily not reject the null hypothesis of *I*(2) even when the data are in fact *I(1)*. Indeed, once we control for these changes, our results contradict the findings in Beenstock et al. (2012).

To focus on the first-derivative CO_2 data, which is relevant to our paper, we note that Pretis and Hendry (2013) show that, when the series are broken up into their two underlying series each measured in its own way and assessed using the ADF procedure, the null hypothesis (that the first-derivative CO_2 series is non-stationary) is rejected. In other words, Pretis and Hendry (2013) find first-derivative atmospheric CO_2 to be stationary (I (1)) as we do.

Comment 3

The present authors report ADF tests which reject unit roots (e.g. Table 3) yet it is clear from Figure 3 that the series exhibit an upward drift – clearly not stationary, although possibly "trend stationary". This would need to be allowed for by including a trend term in the statistic and using the appropriate Dickey-Fuller table. Otherwise, these ADF results are not valid. This issue of the treatment of drift has not been discussed anywhere that I can see, but it definitely needs to be.

Our ADF tests included an allowance for drift and trend in the underlying regressions, and we should have stated this explicitly. We suggest the following changes to the text:

1. Table 3 - Amend the Table heading: Augmented Dickey–Fuller (ADF) tests for unit roots in monthly data ....etc.

Put an asterisk on the column heading ADF statistic*

Then add a footnote to the table: * The Dickey-Fuller regressions allowed for both drift and trend; the augmentation level was chosen by minimizing the Schwarz Information Criterion.

2. Page 29117, starting at line 7: ....Dickey–Fuller (ADF) test for unit roots Table 3 provides the information concerning the stationarity for the level of, and first-derivative of, CO_2, as well as global surface temperature. The test was applied with an allowance for both a drift and deterministic trend in the data, and the degree of augmentation in the Dickey-Fuller regressions was determined by minimizing the Schwarz Information Criterion.

Comment 4

In page 29109 line 11 the authors say "temperature is not stationary of itself but must be made stationary by differencing . . ." (my emphasis). It is important to make clear, something on which the authors are at best equivocal , that a time series cannot be made stationary. It either is stationary, or it isn't. The differences of a series are a different series! It is not difficult to construct examples where the sign of the relationship between two series is reversed in their differences, or where two series are correlated in differences by exhibit independent stochastic trends. Since the AGW hypothesis is that more CO_2 in the atmosphere translates into higher surface temperatures (not that temperatures respond to changes, but not to levels), this fact is crucial in understanding the results of this study. They really don't receive sufficient discussion here. Are these results viewed as supportive of the AGW hypothesis, or not? We would appear to need continuously accelerating growth in CO_2 *to produce warming on an alarming scale. Is this hypothesis proposed, and what mechanism is envisaged? These questions badly need answering, or at least posing, if the reported results are to be understood.*

We will deal with the elements of this Comment in the following order:

…what mechanism is envisaged?

Referring to "mechanism" in the sense widely used in science (for example, Machamer et al. (2000): *an entity and activity productive of regular changes in a separate entity*), we nominated as the candidate entity the terrestrial biosphere. This has already been widely proposed in climate science. For example, from page 29104:

> "It is widely considered that the interannual variability in the growth rate of atmospheric CO_2 is a sign of the operation of the influence of the planetary biota. Again, IPCC (2007) states: "The atmospheric CO_2 growth rate exhibits large interannual variations. The change in fossil fuel emissions and the estimated variability in net CO_2 uptake of the oceans are too small to account for this signal, which must be caused by year-to-year fluctuations in land–atmosphere fluxes." In the IPCC Fourth Assessment Report, Denman et al. (2007) state (italics denote present author emphasis): "Interannual and inter-decadal variability in the growth rate of atmospheric CO_2 is dominated by the *response of the land biosphere to climate variations*. . . . The terrestrial biosphere *interacts strongly with the climate*, providing both positive and negative feedbacks due to biogeophysical and biogeochemical processes. . . . Surface climate is determined by the balance of fluxes, which can be changed by radiative (e.g., albedo) or non-radiative (e.g., water cycle related processes) terms. Both radiative and non-radiative terms *are controlled by details of vegetation*."

In Machamer et al. 2000 terms, we have provided evidence that the terrestrial biosphere is a candidate mechanism for the climate effects as follows: the evidence (by correlation) is that the *entity* of the terrestrial biosphere contains *activities* – depicted by the NDVI time series – which are *productive of regular changes,* as seen in the *separate entity* of the atmosphere.

The point being raised, we have attempted to utilise the concept of mechanism more widely to sharpen our description of the other climate influences discussed in the paper.

We suggest therefore the following series of amendments or additions to the text (shown in italics) to more clearly utilise the concept of mechanism:

Page 29103, Line 19:

The situation is illustrated visually in Fig. 1 which shows the increasing departure over recent years of the global surface temperature trend from that projected by a representative climate model (the CMIP3, SRESA1B scenario model for global surface temperature, KNMI 2013). It is noted that the level of atmospheric CO_2 is a good proxy for the IPCC models predicting the global surface temperature trend: according to IPCC AR5 (2013), on decadal to interdecadal time scales and under continually increasing effective radiative forcing, the forced component of the global surface temperature trend responds to the forcing trend relatively rapidly and almost linearly. *This trend can be taken to represent that expected from the operation of the standard anthropogenic global warming model, its mechanism being a physical one in which (IPCC, 2013, NASA 2015) about half the light reaching Earth's atmosphere passes through the air and clouds to the surface, where it is absorbed and then radiated upward in the form of infrared heat. About 90 percent of this heat is then absorbed by the greenhouse gases and radiated back toward the surface, which is warmed. If greenhouse gases have been increasing (including because of increasing anthropogenic emissions), that contributes to an increase in the infrared radiation they emit (including that back toward the surface, which is warmed further).*

Page 29104, Line 5:

A wide range of physical explanations has now been proposed for the global warming slowdown. These involve proposals either for changes in the way the *radiative mechanism itself* is working or for the increased influence of *other physical mechanisms*. Chen and Tang (2014) place these proposed explanations into two categories. The first involves a reduction in radiative forcing: by a decrease in stratospheric water vapour, an increase in background stratospheric volcanic aerosols, by 17 small volcano eruptions since 1999, increasing coal-burning in China, the indirect effect of time-varying anthropogenic aerosols, a low solar minimum, or a combination of these. The second category of candidate explanation involves planetary sinks for the excess heat. The major focus for the source of this sink has been physical and has involved ocean heat sequestration. However, evidence for the precise nature of the ocean sinks is not yet converging: according to Chen and Tang (2014) their study followed the original proposal of Meehl et al. (2011) that global deep-ocean heat sequestration is centred on the Pacific. However, their observational results were that such deep-ocean heat sequestration is mainly occurring in the Atlantic and the Southern oceans.

Alongside the foregoing possible physical causes, Hansen et al. (2013) have suggested that *the mechanism for* the pause in the global temperature increase since 1998 might be the planetary biota, in particular the terrestrial biosphere.

Page 29124, Line 23:

4.4 Normalized Difference Vegetation Index (NDVI) data

This section now investigates the land biosphere as a candidate *mechanism* for the foregoing effects, in particular the increasing difference between the global surface temperature trend suggested by general circulation climate models and that observed.

Page 29127, Line 3:

A second notable finding highlighted by the bringing together of results in Table 12 is the major role of immediate past instances of the dependent variable in its own present state. This was found to be the case in all the instances where time series models could be prepared. This was true for both temperature and SOI. This was not to take away from first and second-derivative CO_2 – in all the cases just mentioned, they were significant in the models as well. Further, and perhaps equally notably, each was shown to be Granger-causal to its relevant climate outcome.

*Turning to the Normalized Difference Vegetation Index analysis***,** *Figure 10 and Table S4 show that the NDVI signature closely fits the difference between the global surface temperature trend suggested by general circulation climate models and that observed. This fit provides evidence that the terrestrial biosphere mechanism is the cause of the departure of temperature from that predicted by the radiative forcing mechanism alone. In other words, Figure 10 provides evidence that of the two mechanisms in operation together. (It is notable that CO_2 is having two different influences on climate through two quite different mechanisms – the first, a radiative one, with CO_2 as a greenhouse gas, the second as a result of plants requiring CO_2 as a resource*.)

Are these results viewed as supportive of the AGW hypothesis, or not?

The results are supportive of the anthropogenic global warming hypothesis that variations in atmospheric carbon dioxide influence surface temperature. First-derivative atmospheric CO_2 is shown to drive global temperature and the results deepen the support for *CO_2* affecting climate in that second-derivative CO_2 is shown to drive the SOI.

Lastly, the results show that the NDVI signature fits the difference between the global surface temperature observed trend and that suggested by the standard AGW hypothesis / radiative forcing mechanism. This fit provides evidence that the terrestrial biosphere mechanism is the cause of this departure of temperature from that predicted by the standard AGW hypothesis / radiative forcing mechanism alone.

The results, then, are supportive of the anthropogenic global warming hypothesis. The proviso is that the results provide evidence that the final warming achieved is the result not of one mechanism – the physical greenhouse gas radiative mechanism embodied in the standard anthropogenic global warming hypothesis - but of the interaction of that mechanism with a second, residing in the terrestrial biosphere.

We suggest therefore the following additions to the text:

Page 29127, after Line 10:

> The results are supportive of the anthropogenic global warming hypothesis that variations in atmospheric carbon dioxide influence surface temperature. First-derivative atmospheric CO_2 is shown to drive global temperature and the results deepen the support for *CO_2* affecting climate in that second-derivative CO_2 is shown to drive the SOI. Lastly, the results show that the NDVI signature fits the difference between the global surface temperature observed trend and that suggested by the standard AGW hypothesis / radiative forcing mechanism. This fit provides evidence that the terrestrial biosphere mechanism is the cause of this departure of temperature from that predicted by the standard AGW hypothesis / radiative forcing mechanism alone. In other words, the results provide evidence for the case that the final warming achieved is the result not of one mechanism – the physical greenhouse gas radiative mechanism embodied in the standard anthropogenic global warming hypothesis - but of the interaction of that mechanism with a second, residing in the terrestrial biosphere.

We would appear to need continuously accelerating growth in CO_2 *to produce warming on an alarming scale. Is this hypothesis proposed… These questions badly need answering, or at least posing…*

As mentioned in the Introduction, the standard notion of the greenhouse effect (IPCC, 2013) has it that global temperature will rise almost linearly with an increasing level of global atmospheric CO_2. We certainly note here that from the NDVI section of the present paper that there has been an increasing NDVI over recent years and that that correlates with global temperature trending below that predicted by the standard AGW hypothesis / radiative forcing mechanism.

Questions which can be posed from these results include those of (i) under what conditions can the current increase in plant biomass be expected to continue, and (ii) what is the range of alternative expected future trajectories for human greenhouse gas emissions? Obviously the combinations of the extremes of these ranges produce quite different climate trend outcomes.

We suggest therefore the following additions to the text:

Page 29127, before Line 11:

> As mentioned in the Introduction, the standard notion of the greenhouse effect (IPCC, 2013) has it that global temperature will rise almost linearly with an increasing level of global atmospheric CO_2. We note here that from the NDVI section of the present paper that there has been an increasing NDVI over recent years and that that correlates with global temperature trending below that predicted by the standard AGW hypothesis / radiative forcing mechanism.
>
> Questions arising from these results include those of (i) under what conditions can the current increase in plant biomass be expected to continue, and (ii) what is the range of alternative expected future trajectories for human greenhouse gas emissions? Obviously the combinations of the extremes of these ranges produce quite different climate trend outcomes.

Comment 5

In their analysis of the monthly data, the authors explain how they have smoothed the CO_2 series by a moving average (Page 29113, line 10). This is evident in any case, because the raw CO_2 series is highly seasonal, and no seasonality is apparent here. The problem is that smoothing and seasonal adjustment filters are notorious for changing the dynamics of relationships. I do not see how the lag-correlograms of Figures 4 and 5 are to be interpreted if they are computed for smoothed and deseasonalised data. They really prove nothing – and the same criticism has to be made of the various Granger causality tests reported, if these are conducted on smoothed data. The only legitimate way to conduct these kind of tests, where timing shifts of one or two months is critical, is on the raw observations, where extraneous data features such as seasonality have been accounted for by effective modelling. This may be tricky, but in the case of a seasonal pattern it might, for example, be effective to employ polynomial dummy variables to explain seasonal changes, We turn first to “The problem is that smoothing and seasonal adjustment filters are notorious for changing the dynamics of relationships.”

We address this point in two ways. The first is to assess empirically with our data sets the extent to which the filters used did cause changes in dynamics. Secondly, we make observations on the literature on this topic.

Assessment 1. Does the smoothed first-derivative CO_2 series used in the paper have different key dynamics compared with the original raw (unsmoothed) data from which the smoothed series was derived?

First we reproduce here Figure 4 and Table 1 from the paper. These illustrate the prime aspects of our assessment of which of first-derivative atmospheric CO_2 and global surface temperature leads which (has priority).

Response Figure 1

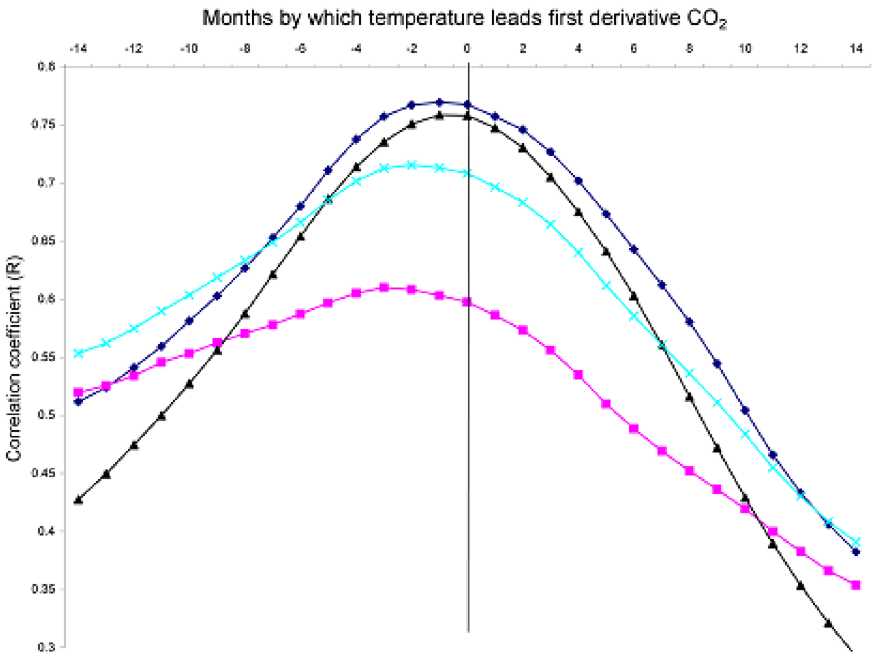


Figure 4. Correlograms of first-derivative CO_2 with surface temperature for global (turquoise curve), tropical (black), Northern Hemisphere (purple) and Southern Hemisphere (blue) categories.

Response table 4

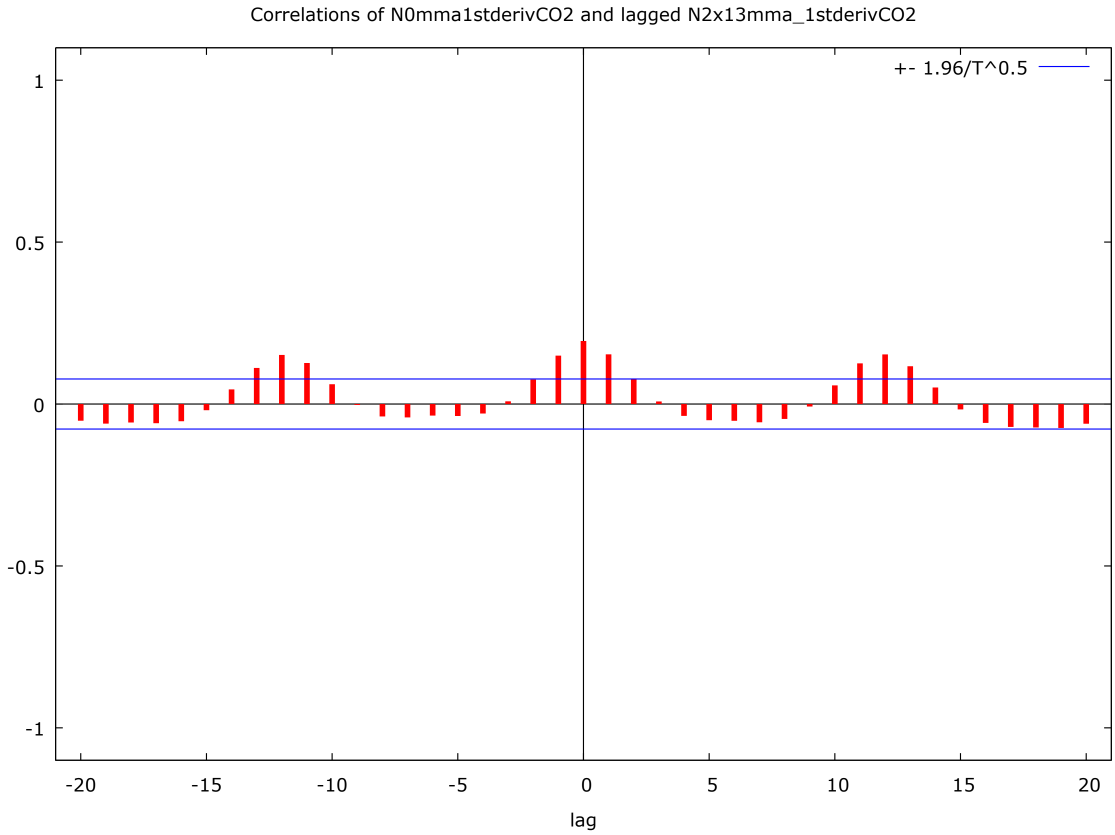
Table 1. Lag of first-derivative CO_2 relative to surface temperature series for global, tropical, Northern Hemisphere and Southern Hemisphere categories.

	Lag in months of first-derivative CO_2 relative to global surface temperature category
hadcrut4SH	−1
hadcrut4Trop	−1
HadCRUT4_nh	−3
hadcrut4Glob	−2

The key point from the above (and the next figure and table in the paper) is that in all cases assessed, first-derivative atmospheric CO_2 led global surface temperature.

In these analyses, only the CO_2 series was smoothed and therefore requires assessment. To do this, let us see if the smoothed first-derivative CO_2 series used in the paper has different key dynamics to that of the original raw (unsmoothed) data from which the smoothed series was derived. Lagged correlogram analysis is used to assess this question. In the tables presented, degree of statistical significance is indicated by stars: one star is $p<0.05$, two stars is $p<0.01$ and three stars is $p<0.001$. In the tables and figures, the notation is the same as described in the paper. The exception is to do with the letter “Z” (for Z score). Here Z is sometimes replaced by “N”. This stands for Normalised, and has the same meaning.

Response Figure 2



Response Table 5

	1stderivCO_2 and 2x13mma1stderivCO_2	Statistical significance
-20	-0.0515	
-19	-0.0605	
-18	-0.0572	
-17	-0.0593	
-16	-0.0532	
-15	-0.0191	
-14	0.0451	
-13	0.1113	***
-12	0.1516	***
-11	0.1267	***
-10	0.0611	
-9	-0.0029	
-8	-0.0383	
-7	-0.0413	
-6	-0.0357	
-5	-0.037	
-4	-0.0293	
-3	0.0083	
-2	0.0753	*
-1	0.1494	***
0	0.1946	***
1	0.1535	***
2	0.0788	**
3	0.0079	
4	-0.0367	
5	-0.05	
6	-0.0518	
7	-0.0563	
8	-0.0461	
9	-0.0078	
10	0.0576	
11	0.1255	***
12	0.1532	***
13	0.1167	***
14	0.051	
15	-0.0167	
16	-0.0583	
17	-0.0707	*
18	-0.0724	*
19	-0.074	*
20	-0.0609	

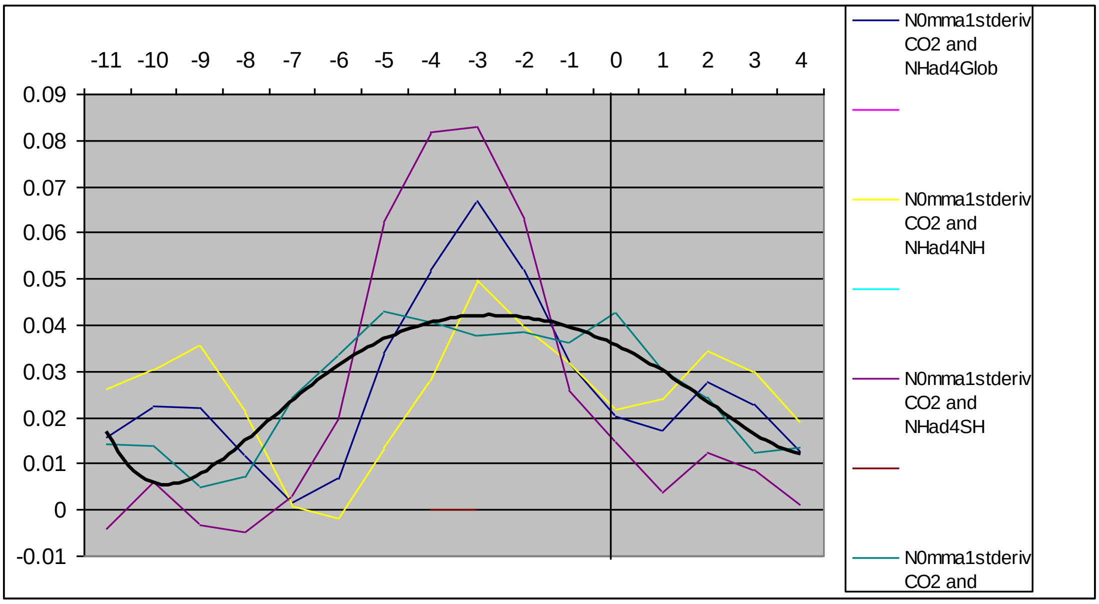
In the figure and the table it can be seen that the maximum, and statistically significant, correlation of the smoothed series with the unsmoothed series is when there is no phase shift. This suggests the particular smoothing used in the paper should provide no problems in the assessment of which of first-derivative CO_2 and temperature has priority. A similar lack of phase problems can be shown between smoothed (2 x 13 month moving average) and unsmoothed second-derivative CO_2 used later in the paper.

That said, we can also carry out a further robustness check by repeating the analysis shown in Figure 4 and Table 1 in the paper (Page 8 above), now using data for the original unsmoothed (raw) first-derivative CO_2 data.

Response Table 6

Lag	0mma1stderivCO₂ and NHad4Glob	Statistical significance	0mma1st derivCO₂ and NHad4NH	Statistical significance	0mma1st derivCO₂ and NHad4SH	Statistical significance	0mma1st derivCO₂ and NHad4Trop	Statistical significance
-11	0.016		0.026		-0.004		0.014	
-10	0.022		0.030		0.006		0.014	
-9	0.022		0.035		-0.003		0.005	
-8	0.012		0.021		-0.005		0.007	
-7	0.002		0.001		0.003		0.024	
-6	0.007		-0.002		0.020		0.034	
-5	0.034		0.014		0.062		0.043	
-4	0.052		0.028		0.082	**	0.041	
-3	0.067	*	0.050		0.083	**	0.038	
-2	0.052		0.040		0.063		0.039	
-1	0.032		0.032		0.026		0.036	
0	0.020		0.022		0.015		0.043	
1	0.017		0.024		0.004		0.030	
2	0.028		0.034		0.012		0.024	
3	0.023		0.030		0.009		0.012	
4	0.013		0.019		0.001		0.014	

Response Figure 3



Response Table 7

	2x13mma (L&B 2015)	0mma1stderivCO_2
NHad4SH	-1	-3
NHad4Trop	-1	-3
NHad4NH	-3	-3
NHad4Glob	-2	-3

It is noted that due to the effect of the seasonality also being present, the correlations in Response Figure 3 are much lower than those from the deseasonalised series used in the paper (Response Figure 1). Nonetheless, the point of the assessment in the paper – to see which of first-difference CO_2 and temperature has priority, and the finding for first-difference CO_2 – is completely confirmed by use of data with no smoothing.

The literature is extensive on the effect that seasonal adjustment has on a number of the assessments carried out in the paper. With regard to the tests for unit roots in time-series data, for example Ghysels (1990), Frances (1991), Ghysels and Perron (1993), Diebold (1993), and Maddala and Kim (1998, pp. 364-365) discuss the fact that *in finite samples* the ADF test is biased towards non-rejection of the unit root null hypothesis if the data are smoothed or filtered to eliminate deterministic seasonality.

That is, their power is reduced. However, this distortion is *not* an issue with large sample sizes. Moreover, Olekalns (1994) shows that seasonal adjustment using frequency domain (rather than time domain) filters, or by using seasonal dummy variables, also impacts adversely on the finite-sample power of the ADF test.

Next we turn to the point that the modelling itself and the Granger causality testing should have been undertaken with raw (rather than smoothed) data.

How does temporal aggregation, *or smoothing*, of the data affect tests for Granger causality?

A number of authors have addressed this question, including Sims (1971), Wei (1982), Christiano and Eichenbaum (1987), Marcellino (1999), Breitung and Swanson (2002), and Gulasekaran and Abeysinghe (2002).

One of the results emerging from this literature is that while Granger causality can be "masked" by the smoothing of the data, apparent causality cannot be "created" from non-causal data.

We believe that this means that our results relating to the existence of Granger causality should not be affected by the smoothing of the data.

Suggested changes to the paper

On page 29113, **add two new paragraphs** between lines 17 and 18:

> "It is important to consider what effects this filtering of our data may have on the ensuing statistical analysis. In these analyses, only the CO_2 series was smoothed and therefore requires assessment. To do this we tested if the smoothed (2 x 13 month moving average) first-derivative CO_2 series used here has different key dynamics to that of the original raw (unsmoothed) data from which the smoothed series was derived. Lagged correlogram analysis showed that the maximum, and statistically significant, correlation of the smoothed series with the unsmoothed series occurs when there is no phase shift. This suggests that the particular smoothing used should provide no problems in the assessment of which of first difference CO_2 and temperature has priority.
>
> Second, there is extensive evidence that while the effect that seasonal adjustment (via smoothing) on the usual tests for unit roots in time-series data is to reduce their power in small samples, this distortion is *not* an issue with samples of the size used in this study. For example, see Ghysels (1990), Frances (1991), Ghysels and Perron (1993), and Diebold (1993). Moreover, Olekalns (1994) shows that seasonal adjustment by using dummy variables also impacts adversely on the finite-sample power of these tests, so there is little to be gained by considering this alternative approach. Finally, one of the results emerging from the Granger causality literature is that while such causality can be "masked" by the smoothing of the data, apparent causality cannot be "created" from non-causal data. For example, see Sims (1971), Wei (1982), Christiano and Eichenbaum (1987), Marcellino (1999), Breitung and Swanson (2002), and Gulasekaran and Abeysinghe (2002). This means that

> our results relating to the existence of Granger causality should not be affected adversely by the smoothing of the data that has been undertaken."

Comment 6

(Page 29019, line 20) The authors are right to avoid autocorrelation corrections in regression. In econometric practice such corrections, sometimes called "Cochrane-Orcutt" methods, are nowadays discredited since they have the potential to distort the relationships of interest. The authors are correct that dynamic modelling is the right technique. They are also correct (but could emphasize this more explicitly) that regression analysis (which I assume is taken to include contemporaneous drivers) cannot test causality, but can at best calibrate an (untestable) assumption of causality. The Granger-style test is the only legitimate means to explore causality in time series. I think the authors appreciate this fact, but their defence of their approach could be more clearly articulated.

To address this, we suggest re-phrasing the paragraph on page 29109, beginning at line 19, as follows:

> Rather than using a formal Granger causality analysis, a number of authors have instead used conventional multiple regression models in attempts to quantify the relative importance of natural and anthropogenic influencing factors on climate outcomes such as global surface temperature. These regression models use contemporaneous explanatory variables. For example, see Lean and Rind (2008, 2009); Foster and Rahmstorf (2011); Kopp and Lean (2011); Zhou and Tung (2013). This type of analysis effectively assumes a causal direction between the variables being modelled. It is incapable of providing a proper basis for testing for the presence or absence of causality. In some cases account has been taken of autocorrelation in the model's errors, but this does not overcome the fundamental weakness of standard multiple regression in this context. In contrast, Granger causality analysis that we adopt in this paper provides a formal testing of both the presence and direction of this causality (Granger, 1969).

Comment 7

(Page 29110, line 2) How can an "anthropogenic warming trend" be an explanatory variable or influencing factor? This seems to seriously beg the question. There are anthropogenic trends (e.g. level of industrial output) and warming trends (rising temperature?) but if we already know that these are one and the same, we need not bother with studies such as this one! I know the authors are commenting on previous studies here, but elucidation would nonetheless be most desirable.

The use of “warming” was an accidental misstatement. We suggest replacing “warming” with “greenhouse gas (the predominant anthropogenic greenhouse gas being CO_2).”

Comment 8

(Page 29114, line 5) A Dickey-Fuller test is not a test of stationarity. It is a test of a unit root, and there are nonstationary cases of the alternative hypothesis. A test of stationarity (as the null hypothesis) might be the KPSS test (Kwiatkowski et al. (1992), Journal of Econometrics 54, 159-178). However, the KPSS test is not strictly a test of stationarity either. It is a test of weak dependence (i.e., summability of the autocovariance sequence) which is not a necessary condition for stationarity, as such, although it is a condition for conventional inference based on correlations to be valid in large samples, via the central limit theorem. Care needs to be taken to distinguish these different time series properties, and the statistical techniques appropriate to them.

To address this we suggest replacing the last sentence in the first paragraph on page 29114 with:

> The (augmented) Dickey-Fuller test is applied to each variable. For this test, the null hypothesis is that the series has a unit root, and hence is non-stationary. The alternative hypothesis is that the series is integrated of order zero.

Comment 9

(Page 29114, line 21) Pankraz (1991). Reference missing.

Reference will be added.

Comment 10

(Page 29118, line 6) Where is Supplementary Table S1? I don’t think that results should be discussed unless they are included in the paper being submitted for publication.

All Supplementary tables currently in the Supplement accessed by the “Discussion Paper” box at the top right of the ACPD main page for the article http://www.atmos-chem-phys-discuss.net/14/29101/2014/acpd-14-29101-2014.html will be brought into the main paper**.**

Comment 11

(Page 29126, line 24) "data not amenable to time series analysis . . ."? This is an odd statement that needs explaining. How correlations can be "visually observed", if they cannot be tested conventionally, is even odder. I suggest this paragraph needs rethinking, and I will also mention that Figure 9 is puzzling, especially the green plot described as "first derivatives". What are the vertical scales here? Have the curves been shifted and units of measurement changed so as to superimpose them. What's the implication of this? (The same query may be asked about other graphs too).

The components of this Comment will be dealt with in turn.

(Page 29126, line 24) "data not amenable to time series analysis . . ."? This is an odd statement that needs explaining. How correlations can be "visually observed", if they cannot be tested conventionally, is even odder. I suggest this paragraph needs rethinking…

We suggest rewriting the paragraph as follows:

> Table 12 and reference to the relevant figures *and their associated text* show that relationships between first and second-derivative CO_2 and climate variables are present for all the time scales studied, that is, including temporal start points situated as long ago as 1500. In the five instances where time series analysis accounting for autocorrelation could be successfully conducted, the results were statistically significant (two tailed test) in four of the five cases, and near significance in the fifth. For the further instances (commencing in 1500) the data was not amenable to time series analysis *due to the strongly smoothed nature of the temperature data making removal of the autocorrelation impossible (See Section 4.3). Nonetheless the scale of the non--corrected correlations observed (see Table 10) were of the same order of magnitude as those of the other instances listed in Table 12 that were able to be corrected for autocorrelation*. Taken as a whole the results clearly suggest that the mechanism observed is long term, and not, for example, a creation of the period of steepest anthropogenic CO_2 emissions increase which commenced in the 1950s (IPCC, 2013).

…I will also mention that Figure 9 is puzzling, especially the green plot described as "first derivatives".

The green plot is first-derivative ice core CO_2: the caption will be re-written to add this text.

What are the vertical scales here? Have the curves been shifted and units of measurement changed so as to superimpose them. What's the implication of this? (The same query may be asked about other graphs too).

In Section 3 Data and methods (page 29110) we wrote:

> To make it easier to visually assess the relationship between the key climate variables, the data were normalised using statistical *Z* scores or standardised deviation scores (expressed as "Relative level" in the figures). In a *Z* scored data series, each data point is part of an overall data series that sums to a zero mean and variance of 1, enabling comparison of data having different native units.

To address this aspect of comment 11 we suggest adding the following, after "units":

> Hence, when several Z-scored time series are depicted in a graph, all the time series will closely superimpose, enabling visual inspection to clearly discern the degree of similarity or dissimilarity between them.

Comment 12

Final comment. Many readers will have the paper as a monochrome print-out, and for such readers the colour-coded graphs cannot be deciphered. BW versions, with patterns instead of colours to distinguish the curves, are a must!

Graphs will be redone including patterns.

References

Breitung, J. and N. R. Swanson, 2002. Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis*, 23, 651-665.

Christriano, L. J. and M. Eichenbaum, 1987. Temporal aggregation and structural inference in macroeconomics. *Carnegie-Rochester Conference Series on Public Policy*, 26, 63-130.

Diebold, F. X., 1993. Discussion: Effect of seasonal adjustment filters on tests for a unit root. *Journal of Econometrics*, 55, 99-103.

Frances, P. H., 1991. Moving average filters and unit roots. *Economics Letter*s, 37, 399-403.

Ghysels, E., 1990. Unit root tests and the statistical pitfalls of seasonal adjustment: The case of U.S. postwar real gross national product. *Journal of Business and Economic Statistics*, 8, 145-152.

Ghysels, E. and P. Perron, 1993. The effect of seasonal adjustment filters on tests for a unit root. *Journal of Econometrics*, 55, 57-98.

Gulasekaran, R. and T. Abeysinghe, 2002. The distortionary effects of temporal aggregation on Granger causality. Working Paper No. 0204, Department of Economics, National University of Singapore.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. Philosophy of Science, 67, 1–25.

Maddala, G. S. and I-M. Kim, 1998. *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press, Cambridge.

Marcellino, M., 1999. Some consequences of temporal aggregation in empirical analysis. *Journal of Business and Economic Statistics*, 17, 129-136.

Olekalns, N., 1994. Testing for unit roots in seasonally adjusted data. *Economics Letters*, 45, 273-279.

Pierse, R. G. and J. Snell, 1995. Temporal aggregation and the power of tests for a unit root. *Journal of Econometrics*, 65, 333-345.

Wei, W. W. S., 1982. The effect of systematic sampling and temporal aggregation on causality – A cautionary note. *Journal of the American Statistical Association*, 77, 316-319.

IPCC, 2013: Annex III: Glossary [Planton, S. (ed.)]. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

NASA 2015 A blanket around the Earth. http://climate.nasa.gov/causes/