Estimating Climate Sensitivity

J.S.Reid and P.J.Nielsen

March 2022

1 Introduction

Climate science is, or should be, a branch of experimental physics depending on observation rather than theoretical preconceptions. It is primarily concerned with the relationships between time series of measurable physical quantities and their long term behaviour. Here we develop a rigorous and useful statistical calculus of time series and apply it to real world data. We look at how both global and regional surface air temperature time series are related to atmospheric carbon dioxide concentration. We reveal geographic patterns of climate sensitivity which, to date, have eluded climate modellers.

2 Estimating Impulse Response and Sensitivity

In the absence of a moving average component, the regression model which summarises the relationship between two times series x and y is given by

$$Y_m = \alpha_0 x_m + \sum_{n=1}^{p} \alpha_n . y_{m-n} + \Xi_m \quad , \quad m = 1, ..., M \tag{1}$$

where the regression coefficients, α_i, and their confidence limits are estimated using Ordinary Least Squares. It is termed an ARX(p) model for 'autoregressive with exogenous variable'.The sequence of residuals, $\{\xi_m\}$, is given by

$$\xi_m = y_m - \left(\hat{\alpha}_0 x_m + \sum_{n=1}^{p} \hat{\alpha}_n . y_{m-n} \right) \quad , \quad m = 1, ..., M \tag{2}$$

where y_m is the sample value or 'realization' of Y_m and $\hat{\alpha}_0$ to $\hat{\alpha}_p$ are the regression coefficients estimated from the data.

The order, p, is found by testing the residuals, $\{\xi_m\}$, for self correlationby means of the Ljung-Box test (Ljung and Box, 1978). The Ljung-Box Q statistic and its probability, P, are calculated for each candidate order, p. The estimated order, $\hat{p}$, is determined as the least number of coefficients for which P is greater than some predetermined confidence level for which it can be assumed the innovation sequence is not self-correlated.

Our best estimate of the relationship between the two time series is given by the convolution

$$\hat{\gamma} * y = \hat{\alpha}_0 x \tag{3}$$

where

$$\hat{\gamma}_0 = 1 \tag{4}$$

and

$$\hat{\gamma}_n = -\hat{\alpha}_n \quad , \quad n = 1, ..., \hat{p} \tag{5}$$

A more useful form of (3) is

$$y = \hat{I} * x \tag{6}$$

where $\hat{I}$ is given by

$$\hat{I} * \hat{\gamma} = \hat{\alpha}_0 \quad (7)$$

and is termed the impulse response. For display purposes and inter-comparison a normalized impulse response, $\Im$, may be used where

$$\Im * \gamma = 1 \quad (8)$$

Like the regression coefficients, I is a property of the system under investigation and $\hat{I}$ is its estimate. Equation (6) describes the output of the system, y, in response to *any* input sequence, x. The sensitivity of the system, S, is defined here as the response at infinity to a unit step function, H_j, so that

$$S = \sum_{k=0}^{\infty} I_k \quad (9)$$

i.e. it is the sum of the terms of the impulse response. It is a random variable on which confidence limits can be placed. The sum of a convolution is equal to the product of the sums of the convoluting factors. Hence

$$S = \alpha_0 / \sum \gamma_n \quad (10)$$

from which $\hat{S}$ can be estimated.

3 Software

The regression model (1) was solved using the *OLS* method of the *statsmodels* software package (Seabold and Perktold, 2010) to give estimates, $\hat{\alpha}_i$, of the regression parameters, α_i, from which estimates of the other parameters were derived. The *fit* method of *OLS* returns a class instance *results* with properties specifying the parameter estimates and methods which allow confidence limits to be placed, not only on the parameters themselves, but on any algebraic function of them. For example, (4), (5) and (10) were used to define a constraint in the form of a null hypothesis, $\mathcal{H}$, viz.:

$$\mathcal{H} : \hat{\alpha}_0 + \hat{S} \sum_{n=1}^{\hat{p}} \hat{\alpha}_n = \hat{S} \quad (11)$$

From (11) a distribution of the probability of estimated sensitivity, $\hat{S}$, was found found using the t-test method of the *results* class where the relevant R-matrix tuple corresponding to (11) was specified as

$$R = ([1, \hat{S}, ..., \hat{S}], \hat{S}) \quad (12)$$

In Table 1, the Ljung-Box parameter, Q, and its probability, P, were calculated using the associated class *statsmodels.stats.diagnostic.*

4 Global Temperature vs ln(CO2)

We estimated the impulse response of the global average temperature anomaly, T, to the logarithm of atmospheric carbon concentration, $\ln(C)$, rather than to the concentration, C, itself Huang and Shahabadi (2014). From this we derived an estimate of sensitivity, S, and its confidence limits using (11) and (12). Note that sensitivity defined here is the response of the endogenous variable to unit step function in the exogenous variable whereas climate sensitivity, S_c, is defined as the response to doubling of CO_2 concentration. Hence

$$S_c = \ln(2)S \tag{13}$$

and similarly for its confidence limits.

Global average temperature anomaly data, T. were taken from the `HadCRUT.4.5.0.0` data set (Morice et al., 2012). Carbon dioxide concentrations, C, were taken from the University of Melbourne Greenhouse Gas Factsheet (Meinshausen et al., 2017).

The ARX method was applied to annual means of global average temperature, T_i, on the logarithm of atmospheric CO_2 concentration, $\ln(C_i)$, for the interval 1850 CE to 2014 CE. Applying Ljung-Box to the residuals given by (2) for ARX(p), p= 0, ... ,5) gives the results shown in Table 1. The probability, P, for the ARX(4) run has a value of 0.2534 indicating that the null hypothesis that the residuals are unselfcorrelated cannot be rejected. Thus the simplest regression relationship between T_i, and $\ln(C_i)$ which unambiguously fits the data is the ARX(4) model, viz.:

$$T_i = \hat{\alpha}_0 \ln(C_i) + \hat{\alpha}_1 T_{i-1} + \hat{\alpha}_2 T_{i-2} + \hat{\alpha}_3 T_{i-3} + \hat{\alpha}_4 T_{i-4} \quad , \quad i = 5, ..., N \tag{14}$$

with the regression coefficient estimates $\{1.161, 0.509, -0.063, 0.057, 0.199\}$.

This was used in the t-test method of the *results* class to determine probability as a function of estimated sensitivity. From (10), (14) and (13) , we estimated global climate sensitivity as, $S_c = 2.7°$C with 95 percent confidence limits of 2.3°C and 3.4°C.

5 Regional Temperature vs ln(CO2)

A netCDF dataset of local mean monthly temperatures for the entire globe was downloadedMorice et al. (2021). Each temperature value was associated with a 5 degree latitude by 5 degree longitude spherical rectangle, there being 36 X 72 such rectangles. Time series of annual averages were computed for each rectangle for comparison with the above-described time series of $\ln(C_i)$ using the ARX(p) method. The Ljung-Box P value was found for candidate values of p in the range $0 < p \leq 5$. The smallest value of p for which

$$P > 0.1 \tag{15}$$

was chosen as the order of the ARX process and (10) solved for S.

Spherical rectangles were eliminated from further processing when the number of years of good data was less than fifty or when no value of P satisfying (15) could be found. In a surprising

Table 1: Ljung-Box parameter, Q, and its probability, P, for five ARX runs of global average temperature vs. the logarithm of CO_2 concentration.

Run	Variables	Q	P
ARX(0)	T_i vs $\ln(C_i)$ only	284.1	0.0000
ARX(1)	T_i vs $\ln(C_i)$, T_{i-1}	49.0	0.0084
ARX(2)	T_i vs $\ln(C_i)$,T_{i-1},T_{i-2}	48.2	0.0074
ARX(3)	T_i vs $\ln(C_i)$,T_{i-1} to T_{i-3}	40.4	0.0358
ARX(4)	T_i vs $\ln(C_i)$,T_{i-1} to T_{i-4}	29.3	0.2534
ARX(5)	T_i vs $\ln(C_i)$,T_{i-1} to T_{i-5}	28.6	0.2367

number of cases p was zero in which case S was simply the first order regression coefficient of temperature on concentration.

Local sensitivity estimates are mapped in Figure 1. For ease of display S values were rounded to the nearest integer.

6 Discussion

Local climate sensitivity has been mapped before by Asinimov (2001) using a similar statistical method. Our map of local sensitivities (Fig 1) and its close resemblance to Asinimov's map demonstrate the viability of the present method. Extreme values tend to lie close together: for example the occurrence of zero values in the North Atlantic and the occurrence of larger values over land. If our method were not working and only generating noise, there would not be such good spatial correlation.

High values of in Northern Canada and Siberia suggest that, in these locations, CO_2 concentration is the predominant factor controlling nocturnal radiative cooling in the absence of clouds. High regional values in central China have been noted by Eagle et al. (2013). They were derived from paleoclimate data and attributed by them to the fundamental role played by planetary-scale atmospheric dynamics. Low values in the North Atlantic suggest that other factors, such as submarine volcanism, play a role in modulating Sea Surface Temperature so masking radiative effects. Two unusually high maritime values of $S_c = 4°C$ off the SE corner of Australia near the island of Tasmania could be due to the intermittent incursion of the warm East Australia Current into this region.

Figure 1 is only the first, tantalizing glimpse of the possibilities presented by applying rigorous statistical methods to climate. The effect of CO_2 concentration on other types of observations, such as precipitation, wind speed and so on, remain to be explored. The strong spatial gradients in climate sensitivity observed in places such as California, imply that mean temperature *gradients* are also affected by climate change and this too must have meteorological implications.

References

Asinimov, O. A. (2001). Predicting patterns of near-surface air temperature using empirical data. *Climatic Change 50*, 297–315.

Eagle, R. A., C. Risi, J. L. Mitchell, J. M. Eiler, U. Seibt, J. D. Neelin, G. Li, and A. K. Tripati (2013). High regional climate sensitivity over continental china constrained by glacial-recent changes in temperature and the hydrological cycle. *Proceedings of the National Academy of Sciences of the United States of America 110*(22), 8813–8818.

Huang, Y. and M. B. Shahabadi (2014). Why logarithmic? a note on thedependence of radiative forcing on gas concentration. *J. Geophys. Res. Atmos. 119*, 13,683–13,689.

Ljung, G. M. and G. E. P. Box (1978). On a measure of a lack of fit in time series models. *Biometrika 65*, 297–303.

Meinshausen, M., E. Vogel, A. Nauels, and K. Lorbacher (2017). Historical greenhouse gas concentrations for climate modelling (cmip6). *Geosci. Model Dev. 10*, 2057–2116. `https://www.climatecollege.unimelb.edu.au/cmip6`.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 dataset. *Journal of Geophysical Research 117*, D08101.

Morice, C. P., J. J. Kennedy, N. A. Rayner, J. P. Winn, E. Hogan, R. E. Killick, R. J. H. Dunn, T. J. Osborn, P. D. Jones, and I. R. Simpson (2021). An updated assessment of near-surface temperature change from 1850: the hadcrut5 data set. *Journal of Geophysical Research: Atmospheres 126*, e2019JD032361.

Reid, J. S. (2022). The regression of atmospheric concentration on carbon dioxide emissions. *Geophysical Research Letters*. Under Review.

Seabold, S. and J. Perktold (2010). statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.

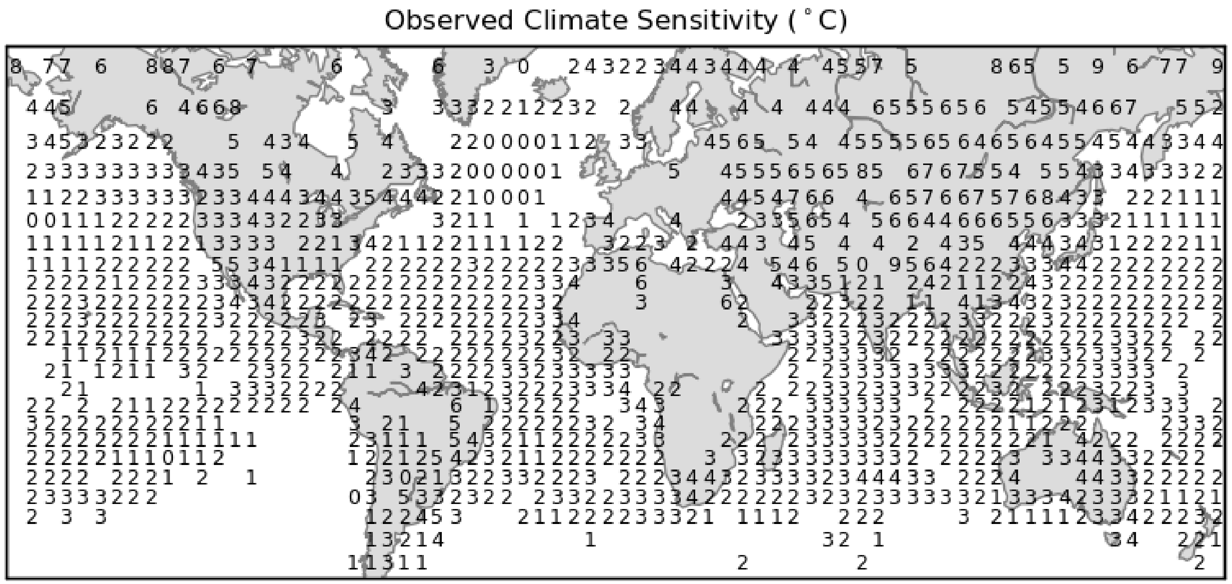

Figure 1: Figure Caption: Climate Sensitivity, S_c estimated for 5 deg X 5 deg rectangles of the earth's surface. Each estimate is derived from the time series of annual average surface air temperature of the rectangle using the ARX method with the time series of ln(CO2 concentration) as the exogenous variable. Only those estimates from rectangles with more than 50 data points and unselfcorrelated residuals are shown. Displayed values have been rounded to the nearest integer.