Statistical Testing of Climate Models

John Reid
johnsimclairreid@gmail.com

Applied Mathematical Sciences

Keywords: Model Testing, Impulse Response, Endogenous Sensitivity, ARX method, Climate Sensitivity, Global Average Temperature, Carbon Dioxide Concentration, Carbon Dioxide Emissions

Statistical Testing of Climate Models

Abstract

Two powerful statistics can be derived from simultaneous time series observations: the impulse response and the endogenous sensitivity. The former is defined as the response to a unit impulse in the exogenous variable and the latter is defined as the sum of terms of the impulse response sequence. These sample statistics provide objective constraints on the output of numerical climate models. Thus climate sensitivity estimated from observed global average temperatures and measured CO_2 concentrations lies in the range 1.9°C to 2.9°C with 95 percent certainty. The estimated impulse response of atmospheric carbon dioxide concentration as a function of emissions is exponential with a half-time of 43 years. The widely accepted hypothesis, that 10 to 20 percent of carbon emissions remain in the atmosphere indefinitely, can be rejected.

Significance Statement

A climate model is a numerical fluid dynamic model in which subsidiary processes involving changes of state and turbulence have been crudely parametrized. It only becomes a scientific theory when it has been exhaustively tested against available observations in accordance with the scientific method practised since the sixteenth century. Real-world time series of global average temperature, atmospheric CO2 concentration and CO2 emissions were compared statistically. Temperature sensitivity was found to be similar to climate model predictions whereas the dependence of concentration on emissions differed radically from model-derived values indicating that concentration would rapidly return to pre-industrial values should anthropogenic emissions cease. Emission reduction measures can therefore be made less stringent than those currently proposed.

Statistical Testing of Climate Models

Introduction

It is common practice across a wide range of sciences to estimate ensemble parameters from sample statistics. A time series is a particularly type of sample, one in which a series of measurements are taken at equal intervals of time or averaged over equal intervals of time. The correlation coefficient is often used to describe the relationship between contemporaneous time series but it is a poor statistic because it does not account for temporal ordering. Two other statistics which better summarize the relationship between two concurrent time series are the impulse response and the endogenous sensitivity. The endogenous sensitivity is defined here as the sum of terms of the impulse response sequence or its integral eqivalent. They can be estimated using the "autoregressive with exogenous variable" or ARX method. The existence of these statistics, and the number of ARX regression coefficients required for their computation, can be established by testing residuals for self-correlation. The impulse response is then found as the convolutional reciprocal of a sequence derived from the ARX regression coefficients. Here we derive these statistics using convolutional methods and apply them to climate time series. These ideas originated with Box and Jenkins (1976) and were recently applied to climate sensitivity by Mills (2019). For continuity and clarity, we reiterate their development here.

Using the ARX Method

For notational convenience, in the following, all sample means have been removed and random variables are assumed to have zero mean.

The autoregressive moving average method with a single exogenous variable, ARMAX(p,q), is given at time, i, by:

$$Y_i = \alpha_0 x_i + \sum_{j=1}^{p} \alpha_j . y_{i-j} + \sum_{k=1}^{q} \beta_j \Xi_{i-k} \quad , \quad i = 1, ..., N \tag{1}$$

where the dependent random variable is Y_i, x_i is the exogenous variable, the y_i are past values of Y_i and the Ξ_i are unselfcorrelated random variables with zero mean. The regression coefficients α_0, α_j and β_j are estimated from the data and p and q are small positive integers. The notation is intended to make a clear distinction between random variables which are Latin upper case, and constants, such as past values of random variables, which are Latin lower case. Equation (1) is a state space representation (Hamilton, 1994) describing states of the system at a succession of discrete instants; the random variable, Y_i, at one instant becomes the constant, y_i, in the following instant. The direction of time is important in regression, which, unlike correlation, allows causality to be inferred.

There are software packages for parameter estimation available under the aegis of the major programming languages. Unfortunately some of these are flawed, because they estimate the exogenous parameter, α_0, prior to estimating the other parameters, leading to omitted-variable bias (Greene, 2003); all parameters must be estimated simultaneously in a regression model. Furthermore, estimation of the MA coefficients, $\{\beta_i\}$, requires an iterative Kalman filter method which may not converge. The second, moving average summation in (1), describes a convoluting or "blurring" function, so that $q > 1$ when the sampling interval, Δt, is too small. Estimation of the MA coefficients can be avoided by decimating the time series by q to give a new time series with a larger sampling interval, $q\Delta t$, for which the innovation sequence, $\{\Xi_m\}$, is unselfcorrelated. Then (1) becomes

$$Y_m = \alpha_0 x_m + \sum_{n=1}^{p} \alpha_n . y_{m-n} + \Xi_m \quad , \quad m = 1, ..., M \tag{2}$$

where $m = qi$, $qM \leq N$, The model summarized by (2) is an ARX(p) model for 'autoregressive with exogenous variable'. The regression coefficients, α_i, and their confidence limits are estimated using Ordinary Least Squares. The sequence of residuals, $\{\xi_m\}$, is given by

$$\xi_m = y_m - \left(\hat{\alpha}_0 x_m + \sum_{n=1}^{p} \hat{\alpha}_n . y_{m-n} \right) \quad , \quad m = 1, ..., M \tag{3}$$

where y_m is the sample value or 'realization' of Y_m and $\hat{\alpha}_0$ to $\hat{\alpha}_p$ are the regression coefficient estimates. The $\{\xi_m\}$ are tested using the Ljung-Box, Q statistic with probability P (Ljung and Box, 1978). The minimum number of coefficients, $\hat{p}$, is found for which P is greater than some confidence level, say 0.1, 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 then

$$\sum_{n=0}^{\hat{p}} \hat{\gamma}_n y_{m-n} = \hat{\alpha}_0 x_m \tag{4}$$

where

$$\hat{\gamma}_0 = 1 \tag{5}$$

$$\hat{\gamma}_n = -\hat{\alpha}_n \quad , \quad n = 1, ..., \hat{p} \tag{6}$$

The sequence $\{\gamma_n\}$ specified by (4) is the prediction error filter of the autoregressive process.

Convolution

We can define a time series more precisely as a finite or semi-infinite sequence, $\{x_0, x_1, x_2, ...\}$ for which the index specifies successive equally spaced intervals of time. The convolution, $c = \{c_k; k = 0, 1, ..., r\} = a*b$, of two time series $a = \{a_i; i = 0, 1, ..., p-1\}$ and $b = \{b_j; j = 0, 1, ..., q\}$, is defined by

$$c_k = \sum_{i+j=k} a_i b_j \tag{7}$$

Under this definition convolution satisfies the commutative, associative and distributive laws of arithmetic. Note also that

$$\sum_i a_i \cdot \sum_j b_j = \sum_k \sum_{i+j=k} a_i b_j = \sum_k c_k \tag{8}$$

Of interest is the commonly occurring case when a time series is a geometric progression, $\{1, \alpha, \alpha^2, \alpha^3, ...\}$. Then

$$\{1, \alpha\} * \{1, \alpha, \alpha^2, \alpha^3, ...\} = \{1\} \tag{9}$$

where $\{1\}$ signifies a time series with a single element, 1, at index zero. Thus $\{1, \alpha\}$ is the convolutional reciprocal of $\{1, \alpha, \alpha^2, \alpha^3, ...\}$.

The sum on the left hand side of (4) is a convolution and (4) can be written

$$\hat{\gamma} * y = \hat{\alpha}_0 x \tag{10}$$

A more useful form of (10) is

$$y = \hat{I} * x \tag{11}$$

where $\hat{I}$ is the convolutional reciprocal of $\hat{\gamma}/\hat{\alpha}_0$ given by

$$\hat{I} * \hat{\gamma} = \hat{\alpha}_0\{1\} \tag{12}$$

and will be termed the estimated impulse response. It can be estimated numerically by iteration using (12) in the form

$$\hat{I}_m = \sum_{i=1}^{p} \hat{\alpha}_i \hat{I}_{m-i} + \hat{\alpha}_0 \delta_m \tag{13}$$

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

$$\Im * \gamma = \{1\} \tag{14}$$

Thus the normalized impulse response is the convolutional inverse of the prediction error filter.

Like the regression coefficients, I is a property of the system under investigation and $\hat{I}$ is its estimate. Equation (11) describes the output of the system, y, in response to *any* input sequence, x.

The endogenous sensitivity of the system, S, is defines as the response at infinity to a unit step function, H_j, where $H_j = 0$ for $j < 0$ and $H_j = 1$ for $j \geq 0$. From (7)

$$S = \lim_{k\to\infty} S_k = \lim_{k\to\infty} \sum_{i+j=k} I_i H_j = \sum_{k=0}^{\infty} I_k \tag{15}$$

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.

According to (8) the sum of a convolution is equal to the product of the sums of the convoluting factors. Thus, from (12)

$$\sum_{m=0}^{\infty} I_m \sum_{n=0}^{p} \gamma_n = S \sum_{n=0}^{p} \gamma_n = \alpha_0 \tag{16}$$

from which $\hat{S}$ can be estimated in terms of the prediction error filter, $\{\hat{\gamma}_n\}$.

Furthermore, from (5), (6) and (16) any given endogenous sensitivity estimate, $\hat{S}$, can be used to define a constraint in the form of a null hypothesis, $\mathcal{H}$, viz.:

$$\mathcal{H} : \hat{\alpha}_0 + \hat{S} \sum_{n=1}^{\hat{p}} \hat{\alpha}_n = \hat{S} \tag{17}$$

From (17) a distribution of the probability of estimated endogenous sensitivity, $\hat{S}$, can be found using the t-test method provided by most Ordinary Least Squares software packages.

Temperature vs ln(CO2)

We estimated the impulse response of the global average temperature anomaly, T, on the logarithm of atmospheric carbon concentration, $\ln(C)$ and from this derived an estimate of climate sensitivity, $\hat{S}_c$, and its probability distribution.

Time series in the form of annual averages of the relevant variables, T and C were downloaded from the Web in November 2020. The 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 Mainshausen et al. (2017) supplemented with recent values from Mauna Loa.

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 (3) 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 0.2301 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.:

$$T_i = \hat{\alpha}_0 \ln(C_i) + \hat{\alpha}_1 T_{i-1} + \hat{\alpha}_2 T_{i-2} + \hat{\alpha}_3 T_{i-3} + \hat{\alpha}_4 T_{i-4} \;\;,\;\; i = 5, ..., N \tag{18}$$

with the regression coefficient estimates $\{1.033, 0.507, -0.064, 0.038, 0.196\}$. The impulse response was estimated using (13) and is shown in Figure 1.

The R-matrix tuple corresponding to (17) is

$$R = ([1, \hat{S}, ..., \hat{S}], \hat{S}) \tag{19}$$

This was used in the t-test method of the *Python statsmodels OLS* software package to determine probability as a function of estimated endogenous sensitivity. The results are shown in Figure 2. Note that the endogenous sensitivity, S, of (16) and (17) is the response to a sustained unit change of the independent variable whereas the climate sensitivity, $\hat{S}_c$, plotted in Figure 2 is conventionally defined as the response to a doubling of atmospheric CO_2 concentration. Hence $\hat{S}_c = \ln(2)\hat{S}$.

CO2 vs Emissions

The response of atmospheric CO2 concentration to an emission pulse in CO2 is discussed extensively by Joos et al. (1994) and, more recently, in the multi-model analysis of Joos et al. (2013). A mathematical formulation was derived numerically by Maier-Reimer and Hasselmann (1987) using a similar model, viz:

$$\Im(t) = A_0 + \sum_{j=1}^{4} A_j \exp(-t/\tau_j) \tag{20}$$

where $\Im(t)$ is the normalized impulse response function and the A_j are the proportions corresponding to various decay times, τ_j as shown Table 2. Equation (20) is plotted as the dashed line in Figure 3. The curve is typical of the IRF curves quoted by Joos *et al.*

For comparison, the ARX method developed here was applied to annual means of atmospheric CO_2 concentration, C_i, for the interval 1850 CE to 2014 CE vs global fossil fuel emissions, E_i. Time series in the form of annual averages of the relevant variables, E, C were downloaded from the Web in November 2020. Global fossil fuel emissions, E, were downloaded from the Carbon Dioxide Information Analysis Center (Boden et al., 2017).

Applying Ljung-Box to the residuals given by (3) for ARX(p), p= 0, ... ,5) resulted in zero probabilities in all cases. Applying the ARMAX method revealed a significant moving average component with $q = 2$. For this reason both time series were decimated by 2 and the ARX / Ljung-Box method reapplied. The results for the decimated data are shown in Table 3. The probability, P, for the ARX(1) run has a value of 0.4359 indicating that the null hypothesis that the residuals are unselfcorrelated cannot be rejected. Thus the simplest regression relationship between C_i and E_i which unambiguously fits the data is the ARX(1) model, viz.:

$$C_i = \hat{\alpha}_0 E_i + \hat{\alpha}_1 C_{i-1} \ \ , \ \ i = 1, ..., N \tag{21}$$

where the estimated regression coefficients are $\hat{\alpha}_0 = 0.21$ and $\hat{\alpha}_1 = 0.969$. The prediction error filter is $\{1, -\hat{\alpha}_1\}$ which has convolutional inverse $\{1, \hat{\alpha}_1, \hat{\alpha}_1^2, ...\}$, a geometric sequence with common ratio $\hat{\alpha}_1$. The nth term of the IRS estimate is given by

$$\hat{I}_n = \hat{\alpha}_0 \hat{\alpha}_1^n = \hat{\alpha}_0 \exp(-\frac{nq\Delta t}{\tau}) \tag{22}$$

and the impulse response can be regarded as discretely sampled from a continuous exponential function with time constant $\tau = -q\Delta t / \ln(\hat{\alpha}_1) = -2/\ln(.969) = 63.5$ years. The half-time is 43 years. The normalized impulse response is shown in Figure 3.

The R-matrix tuple corresponding to (17) is

$$R = ([1, \hat{S}], \hat{S}) \tag{23}$$

This was used in the t-test method of the *Python statsmodels OLS* software package to determine probability as a function of estimated endogenous sensitivity. The results are shown in Figure 4.

Discussion

The climate sensitivity estimate and its confidence limits shown in Figure 2 are not dissimilar from values proposed by the IPCC. On the other hand, the impulse response and endogenous sensitivity of CO_2 concentration estimated here are quite different from conventionally accepted values. The impulse response has an exponential decay with a single time constant (Figure 3) and the endogenous sensitivity estimate is finite (Figure 4). These statistics were derived from an ARX regression model which was an excellent fit to the observations with no self-correlation evident in the residuals (Table 3).

The impulse response function (20) due to Meier-Reimer and Hasselmann must be integrated from zero to infinity with respect to time in order to give the endogenous sensitivity. Their endogenous sensitivity is therefore infinite because of the non-zero term, A_0, in (20) according to which thirteen percent of CO_2 emissions remains in the atmosphere indefinitely. In contrast the endogenous sensitivity estimated here from observations is finite with finite 95 percent confidence limits. This reveals a serious shortcoming of their dynamical global ocean circulation model.

A possible explanation is as follows. The deep ocean is bounded by a turbulent mixed layer and by the highly turbulent Antarctic Circumpolar Current. It is therefore likely to be internally mixed by a Kolmogorov cascade of turbulent eddies, some with spatial scales as large as ocean basins and with time scales of, perhaps, decades. Turbulence is a stochastic phenomenon which is difficult to observe at large spatial and temporal scales and which cannot be readily emulated by deterministic models such as climate models. Eddy diffusion generated by turbulent mixing would greatly increase the capacity of the deep ocean to absorb carbon dioxide and so could account for the shorter half time of the observed impulse response of atmospheric CO_2 concentration.

Whatever the explanation, the long half times and the remnant component of atmospheric CO_2 concentration, presently assumed by most modellers, are not supported by observation.

References

Boden, T. A., G. Marland, and R. J. Andres (2017). Global, regional, and national fossil-fuel co2 emissions. *Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.*

Box, G. E. P. and G. Jenkins (1976). *Time Series Analysis: Forecasting and Control.* San Francisco, CA: Holden-Day.

Greene, W. H. (2003). *Econometric Analysis (5th ed.).* New Jersey: Prentice Hall.

Hamilton, E. J. (1994). *Time Series Analysis.* Princeton, New Jersey: Princeton University Press.

Joos, F., M. Bruno, R. Fink, U. Siegenthaler, T. F. Stocker, C. Le Quéré, and J. L. Sarmiento (1994). An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake. *Tellus B: Chemical and Physical Meteorology 48*(3), 394–417.

Joos, F., R. Roth, J. S. Fuglestvedt, G. P. Peters, I. G. Enting, W. von Bloh, V. Brovkin, E. J. Burke, M. Eby, N. R. Edwards, T. Friedrich, T. L. Frölicher, P. R. Halloran, P. B. Holden, C. Jones, T. Kleinen, F. T. Mackenzie, K. Matsumoto, M. Meinshausen, G.-K. Plattner, A. Reisinger, J. Segschneider, G. Shaffer, M. Steinacher, K. Strassmann, K. Tanaka, A. Timmermann, and A. J. Weaver (2013). Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: a multi-model analysis. *Atmospheric Chemistry and Physics 13*(5), 2793–2825.

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.

Maier-Reimer, E. and K. Hasselmann (1987). Transport and storage of co2 in the ocean - an inorganic ocean-circulation carbon cycle model. *Climate Dynamics 2*, 63–90.

Mainshausen, M., E. Vogel, and N. a. K. L. A (2017). Historical greenhouse gas concentrations for climate modelling (cmip6). *Geosci. Model Dev. 10*, 2057–2116.

Mills, T. (2019). *Applied Time Series Analysis: A Practical Guide to Modelling and Forecasting.* Cambridge, Mass: Academic Press.

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 online. *J. Geophys. Res. 117*, D08101.

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	280.6	0.0000
ARX(1)	T_i vs $\ln(C_i)$, T_{i-1}	46.8	0.0144
ARX(2)	T_i vs $\ln(C_i)$,T_{i-1},T_{i-2}	46.4	0.0115
ARX(3)	T_i vs $\ln(C_i)$,T_{i-1} to T_{i-3}	36.4	0.0841
ARX(4)	T_i vs $\ln(C_i)$,T_{i-1} to T_{i-4}	29.8	0.2301
ARX(5)	T_i vs $\ln(C_i)$,T_{i-1} to T_{i-5}	29.7	0.1964

Table 2: Computed impulse response function parameters in (20) for an increase of the initial atmospheric pCO2 by a factor of 1.25. τ_1, τ_2, τ_3 and τ_4 are the time-constants in years (from Maier-Reimer and Hasselmann (1987)).

A_0	A_1	τ_1	A_2	τ_2	A_3	τ_3	A_4	τ_4
0.131	0.201	362.9	0.321	73.6	0.249	17.3	0.098	1.9

Table 3: Ljung-Box parameter, Q, and its probability, P, for five ARX runs of CO_2 concentration, C, vs. global fossil fuel emissions, E. Both time series were decimated by 2.

Run	Q	pvalue
C(t) vs E(t) only	513.5	0.0000
C(t) vs E(t), C(t-1)	28.5	0.4359
C(t) vs E(t),C(t-1),C(t-2)	28.6	0.3830
C(t) vs E(t) to C(t-3)	24.5	0.5483
C(t) vs E(t) to C(t-4))	24.3	0.5049
C(t) vs E(t) to C(t-5)	22.0	0.5796

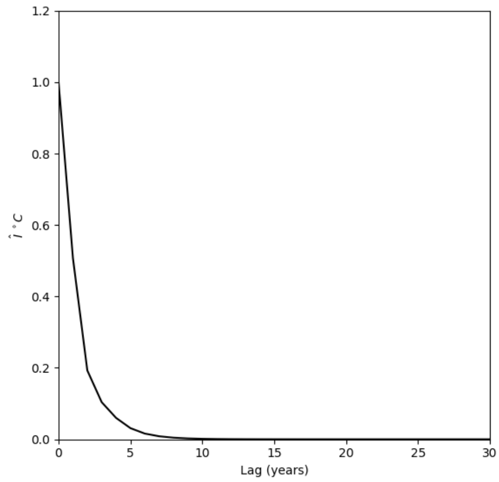

Figure 1: The estimated impulse response, $\hat{I}$, of Global Average Temperature due to a unit impulse in the logarithm of Carbon Dioxide Concentration.

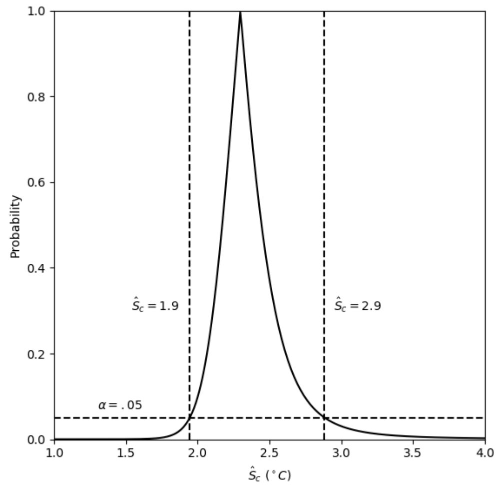

Figure 2: The t-test probablity, P, that the climate sensitivity, S_c, will have a given value. The vertical dashed lines are the 95 percent confidence limits

.

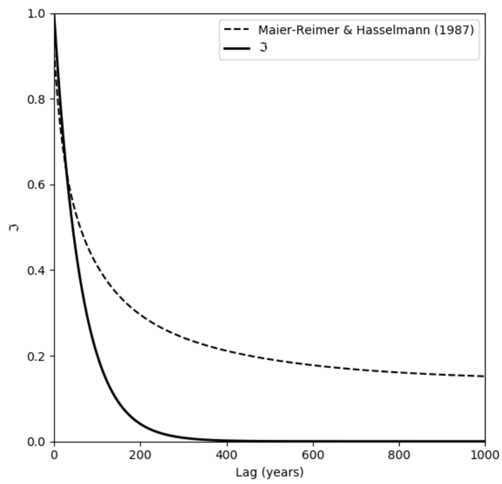

Figure 3: The normalized impulse response, $\Im$, of Carbon Dioxide concentration due to a unit impulse in CO2 Emissions estimated from observations (solid line). Also shown is the model-derived normalized impulse response function of Maier-Reimer and Hasselmann (1987) (dashed line).

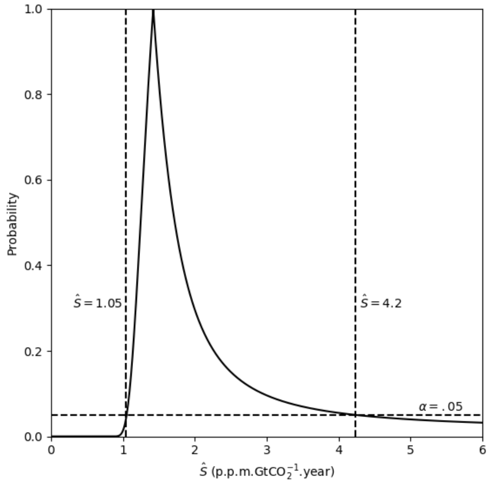

Figure 4: The t-test probablity, P, that the Endogenous Sensitivity, S, CO2 Concentration Sensitivity will have a given value. The vertical dashed lines are the 95 percent confidence limits.

.