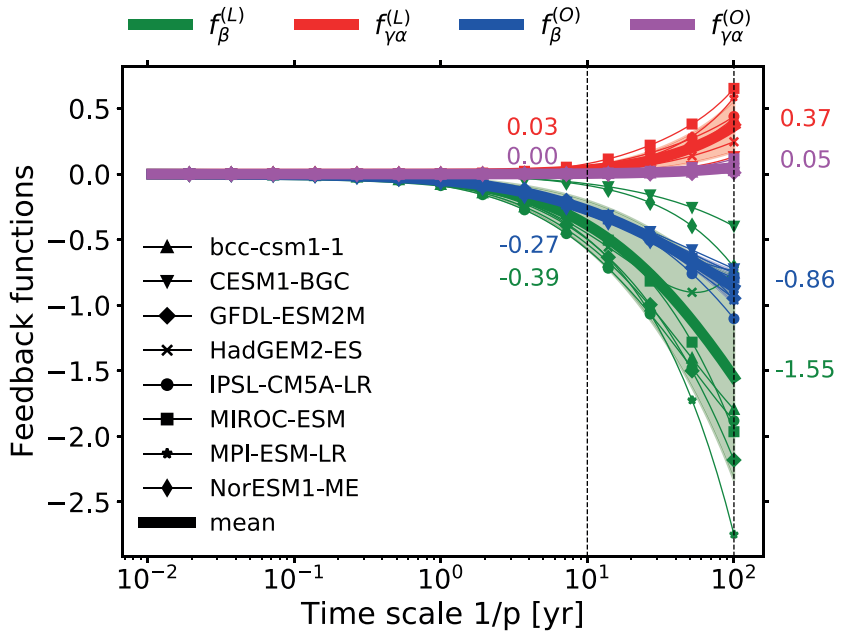
Max-Planck-Institut
für Meteorologie

MAX PLANCK
GESELLSCHAFT

International Max Planck Research School
on Earth System Modelling

Response Characteristics of the Global Carbon Cycle in Earth System Models



Guilherme Luiz Torres Mendonça

Hamburg 2020

Berichte zur Erdsystemforschung
Reports on Earth System Science

237
2020

Hinweis

Die Berichte zur Erdsystemforschung werden vom Max-Planck-Institut für Meteorologie in Hamburg in unregelmäßiger Abfolge herausgegeben.

Sie enthalten wissenschaftliche und technische Beiträge, inklusive Dissertationen.

Die Beiträge geben nicht notwendigerweise die Auffassung des Instituts wieder.

Die "Berichte zur Erdsystemforschung" führen die vorherigen Reihen "Reports" und "Examensarbeiten" weiter.

Notice

The Reports on Earth System Science are published by the Max Planck Institute for Meteorology in Hamburg. They appear in irregular intervals.

They contain scientific and technical contributions, including Ph. D. theses.

The Reports do not necessarily reflect the opinion of the Institute.

The "Reports on Earth System Science" continue the former "Reports" and "Examensarbeiten" of the Max Planck Institute.

Anschrift / Address

Max-Planck-Institut für Meteorologie
Bundesstrasse 53
20146 Hamburg
Deutschland

Tel./Phone: +49 (0)40 4 11 73 - 0
Fax: +49 (0)40 4 11 73 - 298

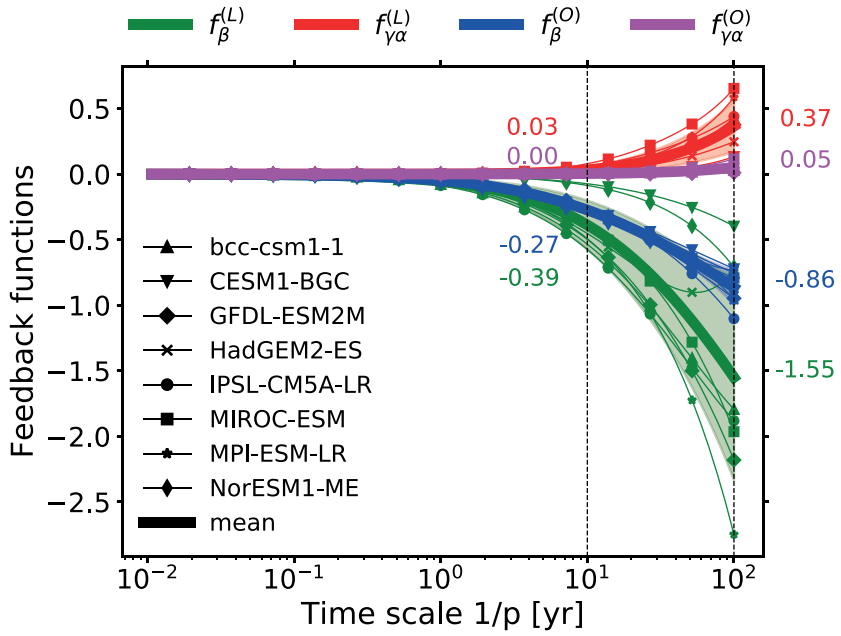
name.surname@mpimet.mpg.de
www.mpimet.mpg.de

Layout

Bettina Diallo and Norbert P. Noreiks
Communication

Copyright

Response Characteristics of the Global Carbon Cycle in Earth System Models



Guilherme Luiz Torres Mendonça

Hamburg 2020

Guilherme Luiz Torres Mendonça
aus Volta Redonda, Brasilien

Max-Planck-Institut für Meteorologie
The International Max Planck Research School on Earth System Modelling
(IMPRS-ESM)
Bundesstrasse 53
20146 Hamburg

Universität Hamburg
Geowissenschaften
Meteorologisches Institut
Bundesstr. 55
20146 Hamburg

Tag der Disputation: 13. Oktober 2020

Folgende Gutachter empfehlen die Annahme der Dissertation:
Dr. Christian H. Reick
Prof. Dr. Martin Claussen

Vorsitzender des Promotionsausschusses:
Prof. Dr. Dirk Gajewski

Dekan der MIN-Fakultät:
Prof. Dr. Heinrich Graener

Berichte zur Erdsystemforschung / Max-Planck-Institut für Meteorologie 237
Reports on Earth System Science / Max Planck Institute for Meteorology 2020

ISSN 1614-1199

Guilherme Luiz Torres Mendonça

Max-Planck-Institut für Meteorologie
Bundestraße 53
20146 Hamburg

"The scientist does not study nature because it is useful to do so. He studies it because he takes pleasure in it, and he takes pleasure in it because it is beautiful. If nature were not beautiful it would not be worth knowing, and life would not be worth living."

– Henri Poincaré

A Isabel e Gilberto, meus maiores exemplos,
e a Gema (1936-2018), minha maior saudade.

ABSTRACT

Typically, the response of the global carbon cycle to atmospheric CO_2 perturbations is characterized by metrics whose values depend on the type of perturbation considered. Here I try to systematically characterize this response in Earth System Models. The approach is to consider weak perturbations and employ the linear response framework. The study focuses especially on the quantification of the climate-carbon cycle feedbacks.

A fundamental step in the application of the linear response framework is to identify the appropriate linear response functions. Existent methods to identify these functions from data require tailored perturbation experiments, which severely limits the applicability of the framework. In the first part of this study, a method is developed to derive linear response functions from arbitrary perturbation experiments. The only requirements are data from a single realization of an experiment and a control simulation. The robustness of the method is tested in several applications using a toy model, including a comparison to existent methods in the literature. I find that under appropriate conditions of signal-to-noise ratio and nonlinearity the presented method generally gives reasonable results, especially if the response function is known to be monotonic. The main novelty of the method consists in the separation of signal and noise by the estimation of the noise level in the data. Due to its generality, this feature may find applications in solving also other types of ill-posed problems.

In the second part of the study, the method is applied to identify in the MPI-ESM (Max Planck Institute for Meteorology Earth System Model) from standard C^4MIP (Coupled Climate Carbon Cycle Model Intercomparison Project) 1% experiments the linear response functions that generalize for weak perturbations the land carbon β- and γ-sensitivities, which characterize the response of land carbon to CO_2 and climate perturbations. In this generalization, the sensitivities are characteristics of the system and are therefore valid for all perturbation scenarios. The robustness of the identified generalized sensitivities is demonstrated by their ability to predict the response of the land carbon in several experiments not used for the identification. The linear regime for which the generalization is valid is estimated. An example demonstrates how response functions recovered with high quality can give useful insight into the internal dynamics of the carbon cycle. For the best recovery of the generalized β-sensitivity, a spectrum of time scales is obtained with peaks at about 4 and 100 years. Robustness of this result is demonstrated by two independent tests. I illustrate how possible explanations for it can be investigated by studying the carbon response for tropics and extra-tropics.

In the third part of the study, a generalization of the climate-carbon feedback framework for weak perturbations is investigated. In this generalization, the feedbacks are described as time-scale dependent characteristics of the system, thus independently of the perturbation.

The generalized framework provides a description of the linear dynamics of the global carbon cycle accounting explicitly for the climate-carbon cycle feedbacks. It is shown that this description predicts with reasonable accuracy the linear dynamics of the carbon cycle in the MPI-ESM. The feedbacks are quantified at different time scales for an ensemble of CMIP5 (Coupled Model Intercomparison Project Phase 5) models. In all models, all feedbacks are zero at small time scales. At larger time scales, the biogeochemical feedbacks are negative and the radiative feedbacks are positive, with the sum of the biogeochemical feedbacks being larger than that of the radiative feedbacks. This results in a negative net feedback at these time scales. From these feedbacks one can as well obtain a time-scale dependent form of the airborne fraction that characterizes the response of atmospheric CO_2 to all emission scenarios. The estimated airborne fraction is 1 for all models at small time scales and decreases to a range of values between 0.26 and 0.5 at a 100-years time scale. The spread in the airborne fraction is largely caused by the spread in the land biogeochemical feedback. This result suggests that research on the carbon cycle should focus especially on improving the understanding of this feedback.

ZUSAMMENFASSUNG

Typischerweise wird die Reaktion des globalen Kohlenstoffkreislaufs auf atmosphärische CO_2-Störungen durch Metriken charakterisiert, deren Werte von der Art der betrachteten Störung abhängen. Hier versuche ich, diese Reaktion in Erdsystemmodellen systematisch zu charakterisieren. Der Ansatz besteht darin, schwache Störungen zu berücksichtigen und den Lineare-Antwort-Formalismus zu verwenden. Die Studie konzentriert sich insbesondere auf die Quantifizierung der Rückkopplungen des Klima-Kohlenstoff-Zyklus.

Ein grundlegender Schritt bei der Anwendung des Lineare-Antwort-Formalismus ist die Identifizierung der geeigneten linearen Antwortfunktionen. Bestehende Methoden zur Identifizierung dieser Funktionen aus Daten erfordern maßgeschneiderte Störexperimente, was die Anwendbarkeit des Formalismus stark einschränkt. Im ersten Teil dieser Studie wird eine Methode entwickelt, um lineare Antwortfunktionen aus willkürlichen Störexperimenten abzuleiten. Die einzigen Anforderungen sind Daten aus einer einzigen Realisierung eines Experiments und aus einer Kontrollsimulation. Die Robustheit der Methode wird in mehreren Anwendungen mit einem Spielzeugmodell getestet, einschließlich eines Vergleichs mit existierenden Methoden in der Literatur. Ich stelle fest, dass die vorgestellte Methode unter geeigneten Bedingungen des Signal-Rausch-Verhältnisses und der Nichtlinearität im Allgemeinen vernünftige Ergebnisse liefert, insbesondere wenn die Antwortfunktion bekanntermaßen monoton ist. Die Hauptneuheit der Methode besteht in der Trennung von Signal und Rauschen durch die Schätzung des Rauschpegels in den Daten. Aufgrund ihrer Allgemeingültigkeit kann dieser Bestandteil auch bei der Lösung anderer Arten von schlecht gestellten Problemen Anwendung finden.

Im zweiten Teil der Studie wird die Methode angewandt, um im MPI-ESM (Max Planck Institute for Meteorology Earth System Model) aus Standard C^4MIP (Coupled Climate Carbon Cycle Model Intercomparison Project) 1%-Experimenten diejenigen linearen Antwortfunktionen zu identifizieren, welche für schwache Störungen die Landkohlenstoff β- und γ-Sensitivitäten verallgemeinern, die die Reaktion des Landkohlenstoffs auf CO_2 und Klimastörungen charakterisieren. In dieser Verallgemeinerung sind die Sensitivitäten Eigenschaften des Systems und daher für alle Störungsszenarien gültig. Die Robustheit der identifizierten verallgemeinerten Sensitivitäten wird durch ihre Fähigkeit zur Vorhersage der Reaktion des Landkohlenstoffs in mehreren Experimenten, die nicht für die Identifizierung verwendet wurden, demonstriert. Das lineare Regime, für das die Verallgemeinerung gültig ist, wird abgeschätzt. Ein Beispiel zeigt, wie mit hoher Qualität wiederhergestellte Antwortfunktionen einen nützlichen Einblick in die interne Dynamik des Kohlenstoffkreislaufs geben können. Für die beste Wiederherstellung der allgemeinen β-Sensitivität wird ein Spektrum von Zeitskalen mit Spitzenwerten bei etwa 4 und 100 Jahren erhalten. Die Robustheit dieses Ergebnisses wird durch zwei unabhängige Tests nachgewiesen. Ich zeige, wie

mögliche Erklärungen dafür untersucht werden können, indem ich die Kohlenstoffreaktion für die Tropen und Extra-Tropen untersuche.

Im dritten Teil der Studie wird eine Verallgemeinerung des Klima-Kohlenstoff-Rückkopplungsformalismus für schwache Störungen untersucht. In dieser Verallgemeinerung werden die Rückkopplungen als zeitskalenabhängige Charakteristika des Systems beschrieben, also als unabhängig von den Störeinflüssen. Der verallgemeinerte Formalismus bietet eine Beschreibung der linearen Dynamik des globalen Kohlenstoffkreislaufs, die explizit die Rückkopplungen des Klima-Kohlenstoffkreislaufs berücksichtigt. Es wird gezeigt, dass diese Beschreibung die lineare Dynamik des Kohlenstoffkreislaufs im MPI-ESM mit angemessener Genauigkeit vorhersagt. Die Rückkopplungen werden auf verschiedenen Zeitskalen für ein Ensemble von CMIP5-Modellen (Coupled Model Intercomparison Project Phase 5) quantifiziert. In allen Modellen sind alle Rückkopplungen auf kleinen Zeitskalen gleich Null. Auf größeren Zeitskalen sind die biogeochemischen Rückkopplungen negativ und die Strahlungsrückkopplungen positiv, wobei die Summe der biogeochemischen Rückkopplungen größer als die der Strahlungsrückkopplungen ist. Dies führt zu einer negativen Netto-Rückkopplung auf diesen Zeitskalen. Aus diesen Rückkopplungen kann man auch eine zeitskalenabhängige Form des CO_2-Anteils in der Luft ableiten, die die Reaktion des atmosphärischen CO_2 in allen Emissionsszenarien charakterisiert. Der geschätzte CO_2-Anteil in der Luft beträgt für alle Modelle auf kleinen Zeitskalen 1 und nimmt auf einer Zeitskala von 100 Jahren auf einen Wertebereich zwischen 0,26 und 0,5 ab. Die Unterschiede des CO_2-Anteils in der Luft werden weitgehend durch die Unterschiede in der biogeochemischen Rückkopplung des Landes verursacht. Dieses Ergebnis legt nahe, dass sich die Forschung über den Kohlenstoffkreislauf besonders auf die Verbesserung des Verständnisses dieser Rückkopplung konzentrieren sollte.

CONTENTS

1 INTRODUCTION 1
- 1.1 The global carbon cycle and climate-carbon cycle feedbacks 2
- 1.2 The α-β-γ framework and its generalization 4
- 1.3 Motivating research questions 5

2 IDENTIFICATION OF LINEAR RESPONSE FUNCTIONS FROM ARBITRARY PERTURBATION EXPERIMENTS 9
- 2.1 Introduction 9
- 2.2 Linear response theory and generalization of climate-carbon cycle sensitivities 12
- 2.3 Identification of linear response functions from arbitrary perturbation experiments 16
  - 2.3.1 Functional form of the linear response function 17
  - 2.3.2 Discretized problem 18
  - 2.3.3 Regularized solution 20
  - 2.3.4 Determining the regularization parameter λ from the noise 22
  - 2.3.5 Additional noise level adjustment in the presence of a monotonicity constraint 24
- 2.4 Applicability in the presence of noise and nonlinearities 25
  - 2.4.1 Toy model and artificial experiments 25
  - 2.4.2 Choice of parameters for the RFI method 28
  - 2.4.3 Ideal conditions 28
  - 2.4.4 First complication: noise 29
  - 2.4.5 Second complication: nonlinearity 33
- 2.5 Comparison with previous methods 40
- 2.6 Summary and discussion 46

3 LINEAR RESPONSE FUNCTIONS AS GENERALIZATION OF LAND CARBON SENSITIVITIES IN THE MPI-ESM 49
- 3.1 Introduction 49
- 3.2 Estimating the linear regime in "percent" experiments 50
- 3.3 Generalized sensitivity χ_γ 52
- 3.4 Generalized sensitivity χ_β 56
- 3.5 Spectrum of land carbon time scales 63
- 3.6 Summary and discussion 71

4 CLIMATE-CARBON CYCLE FEEDBACKS FOR WEAK PERTURBATIONS IN CMIP5 MODELS 75
- 4.1 Introduction 75
- 4.2 Generalized α-β-γ framework 76

4.3 Generalized sensitivities for ocean carbon and temperature in the MPI-ESM 80
4.3.1 Generalized sensitivity $\chi_{\beta}^{(O)}$ 81
4.3.2 Generalized sensitivity $\chi_{\gamma}^{(O)}$ 86
4.3.3 Generalized sensitivities $\chi_{\alpha}^{(L)}$ and $\chi_{\alpha}^{(O)}$ 87
4.4 Airborne fraction in the MPI-ESM determined from the generalized framework 93
4.4.1 Determining the airborne fraction from emission-driven simulations 95
4.4.2 Airborne fraction determined from the generalized framework 97
4.5 Climate-carbon cycle feedbacks for weak perturbations in CMIP5 models 99
4.5.1 Generalized sensitivities 100
4.5.2 Climate-carbon cycle feedbacks and airborne fraction 103
4.6 Summary 110
4.7 Discussion 113
5 CONCLUSIONS 117
5.1 Identification of linear response functions from arbitrary perturbation experiments 117
5.2 Linear response functions as generalization of land carbon sensitivities in the MPI-ESM 118
5.3 Climate-carbon cycle feedbacks for weak perturbations in CMIP5 models 119
5.4 Outlook 121
5.4.1 A critical assessment of the results 121
5.4.2 Implications 123
A APPENDIX TO CHAPTER 2 127
A.1 The Max Planck Institute for Meteorology Earth System Model 127
A.2 Sensitivity of the recovered response function and spectrum to the parameters M, $\log \tau_{min}$ and $\log \tau_{max}$ of the RFI algorithm 127
A.3 Derivation of eqs. (2.14) and (2.15) on which the study is based 133
A.4 Monotonicity of $\chi_{\beta}(t)$ 135
A.5 Spectrum $q(\tau)$ positive or negative for all τ implies $\chi(t)$ monotonic 136
A.6 Response function and noise in the nonlinearized response for the toy model 136
B APPENDIX TO CHAPTER 3 139
B.1 Derivation of spectrum and $\chi(t)$ when the response contains time scales much longer or much shorter than the time scales covered by data 139
B.2 Recovering a discrete spectrum $q(\tau)$ 141
C APPENDIX TO CHAPTER 4 147
C.1 Monotonicity of $\chi_{\gamma}^{(O)}$ in the MPI-ESM 147
C.2 Method to calculate χ_{ζ} 150

C.3 $\chi_\zeta(t)$ non-negative and monotonic implies $0 \leq \widetilde{A}(p) \leq 1$ and interpretation of A 151

BIBLIOGRAPHY 155

Figure 2.1 Forcings for the CMIP-type experiments considered in this study. 13
Figure 2.2 Final RFI algorithm (see text for details). 26
Figure 2.3 Demonstration of robust recovery for noise-free data from the toy model: (a) recovered $\mathbf{q}_\lambda$; (b) recovered $\chi(t)$; and (c) original responses (dashed lines) and predictions (continuous lines) using $\chi_\lambda(t)$ derived from the 1% experiment. Reconstructed values are almost indistinguishable from original data. For plotting the "true" spectrum of the toy model in subfigure (a) I used the relation $\mathbf{q}^* = \mathbf{a}/\Delta \log_{10} \tau$, which can be obtained by comparing eq. (2.36) with eq. (2.15). Since from the discrete spectrum the response function and the response may be obtained for any time t, the spectrum is plotted as dots while the response function and response are plotted as continuous lines. 30
Figure 2.4 Mean prediction error (2.40) of the recovery when deriving $\chi(t)$ for different values of the SNR. As the SNR increases, the recovery of $\chi(t)$ improves. To illustrate the most general case where $\chi(t)$ is not known to be monotonic I do not apply the monotonicity check (step 6 of Fig. 2.2). 32
Figure 2.5 Demonstration of the operation of the RFI algorithm in the presence of noise using toy model data from a 1% and a control experiment. To demonstrate the relevance of the noise level adjustment (step 3 from Fig. 2.2), the standard deviation of the noise in the control simulation was taken ten times smaller than that for the noise in the perturbed simulation. (a) Picard plot showing the singular values σ_i and the projection coefficients of the data $|\mathbf{u}_i \bullet \boldsymbol{\Delta}\mathbf{Y}|$, the "true" noise $|\mathbf{u}_i \bullet \boldsymbol{\eta}|$, and the final noise estimate $|\mathbf{u}_i \bullet \boldsymbol{\eta}_{est}|$; (b) coefficients of regularized solution (2.23); (c) "true" and recovered linear response functions. Since the RFI algorithm correctly adjusted the noise level to the "true" noise in the data, the resulting regularized solution has contributions only from the first few projection coefficients which are not completely obscured by noise. Overall, the recovery is almost perfect, because the SNR (chosen as about 650) is still sufficiently good and because the noise was chosen to conform with the spectral similarity assumption. Because the noise level adjustment (step 3 from Fig. 2.2) already gave a good estimate to the "true" noise in the data, no monotonicity check was needed (step 6 from Fig. 2.2). 34

Figure 2.6 Demonstration of the additional noise level adjustment in the presence of a monotonicity constraint using toy model data from a 1% and a control experiment: (a) Picard plot; (b) coefficients of regularized solution (2.23) and (c) recovered linear response function. All figures are based on the same toy model simulations using a SNR = 1189. To demonstrate the effect of the noise level adjustment the spectral similarity assumption is broken by artificially increasing the low-frequency components of the noise in the simulations. The plots in the first row show the results from the RFI algorithm in the absence of additional noise level adjustment (step 6 in Fig. 2.2). Although the "true" response function of the toy model is monotonic, the response function recovered by the RFI algorithm is non-monotonic (last figure in the first row). But if the noise adjustment is switched on (second row), the response function is correctly recovered as monotonic (last figure in the second row). For more details see text. 35

Figure 2.7 Mean prediction error (2.40) of the recovery when deriving $\chi_\lambda(t)$ for different values of the nonlinearity factor a of the toy model. As a increases, the recovery of $\chi_\lambda(t)$ deteriorates because the level of the contributions from nonlinearities $||\widetilde{\boldsymbol{\eta}}||$ gets large compared to the noise level $||\boldsymbol{\eta}||$; how these terms are computed for the toy model is explained in Appendix A.6. To demonstrate here the pure effect from the breakdown of the spectral similarity assumption, the RFI algorithm is used here without the additional noise level adjustment enforcing monotonicity. 37

Figure 2.8 Demonstration of how nonlinearities affect the recovery of the response function: (a) Picard plot; (b) coefficients of regularized solution (2.23) and (c) recovered linear response function. First row: Nonlinearity factor $a = 10^{-10}$ (no monotonicity check); Second row: Nonlinearity factor $a = 2.5 \times 10^{-4}$ (no monotonicity check); Third row: Nonlinearity factor $a = 2.5 \times 10^{-4}$ (with monotonicity check). The noise is overestimated in the low-frequency spectrum in the third row because nonlinearities yield a derived $\chi_\lambda(t)$ that does not obey the monotonicity constraint. As a consequence, the method increases the level of low-frequency components until the monotonicity constraint is obeyed. The failure to obey the monotonicity constraint and consequent large overestimation of noise in this case can be taken as an indication of the presence of nonlinearities in the response. Note that the "true" linear response function in this nonlinear case $a \neq 0$ is obtained analytically from the linear case $a = 0$ via eq. (2.44) (see Appendix A.6). For more details see text. 39

Figure 2.9 Quality of response function recovery by the full RFI method (including step 6 in Fig. 2.2) in comparison to the pulse and step method. Subscripts at "RFI" indicate the experiment from which the response function was recovered with the RFI method. First row: taking the average over the whole ensemble of toy model experiments for recovery; second row: Performing the recovery for each ensemble member separately. (a_1) Recovered response function; (b_1) recovery error; (c_1) prediction error (2.39); (a_2) example of recovered response function from one ensemble member; (b_2) statistics of recovery error; (c_2) statistics of prediction error (2.39). The prediction error is separately computed for the 0.5% and 0.75% experiments. Taking the ensemble average, all methods perform well (see first row). But taking only one ensemble member, the RFI algorithm gives better recovery and prediction errors than the pulse and step methods when comparing the same responses (see second row). 42

Figure 2.10 Quality of response function recovery by our RFI method excluding step 6 in Fig. 2.2 in comparison to the pulse and step method. Response function is recovered taking the individual response for each ensemble member. Subscripts at "RFI" indicate the experiment from which the response function was recovered with the RFI method. (a) Statistics of the recovery error; (b) example of poor recovery with the RFI algorithm; (c) statistics of the prediction error (2.39); (d) statistics of the recovery error excluding for the $RFI_{1\%}$ the 6.5% of the recoveries with recovery error greater than 1. Once again, the RFI method gives better recovery and prediction errors than the Pulse and Step methods for the same responses. Without accounting for monotonicity the variability in the quality of the recoveries from the 1% experiment increases substantially, but poor recoveries are obtained only in few cases. 45

Figure 3.1 Toy model example for the identification of the linear regime by using additional experiments. Shown is the prediction error (2.39) for the response of 0.5% and 0.75% experiments as obtained from the response function calculated by the RFI method from 1% experiments. The prediction errors are plotted against the final forcing strengths of a sequence of 1% experiments with increasing time series length. The crosses at the minima indicate the final forcing strength for which the response function is optimally recovered (see text). Subfigure (a) shows the behaviour for the fully linear toy model ($a = 0$) and subfigure (b) the behaviour in the presence of nonlinear contributions to the response ($a = 5 \times 10^{-5}$). For the purpose of demonstrating more clearly the increase in the prediction error for a decrease in the forcing strength, I include in the plot cases where the forcing strength is extremely small, corresponding to very small time series lengths. To deal with such cases, I set for a number of data points $N < 30$ the number of time scales $M = N$. For such small number of time scales, usually no plateau in the singular values spectrum is found (step 2 of Fig. 2.2). Therefore, for these special cases I also modify the algorithm to interpret the two smallest singular values as a plateau, since their small magnitude makes them have a similar effect to those singular values belonging to the plateau itself. In addition, to illustrate the most general case where $\chi(t)$ is not known to be monotonic, I exclude here the monotonicity check (step 6 of Fig. 2.2). 51

Figure 3.2 Prediction error (2.39) for the 0.5% and 0.75% rad experiments using $\chi_\gamma(t)$ recovered from the 1% rad experiment. The error is shown as function of the final forcing strength used for the recovery of $\chi_\gamma(t)$. No clear minimum is found so that the recovery seems not to be limited by nonlinearities. 53

Figure 3.3 Forcing temperature $\Delta T(t)$ for 1% and $2\times CO_2$ rad experiments. 54

Figure 3.4 χ_γ derived from 1% and $2\times CO_2$ rad experiments and prediction of model responses using these χ_γ (continuous lines are predictions and dashed lines are responses from the MPI-ESM). Circles indicate the maximum value for which 1.5% and 2% responses are predictable according to the estimate of the linear regime (see text). At the right of subfigures (b) and (c) the prediction error (see eq. (2.39)) is indicated for the different experiments, calculated for the 1.5% and 2% rad experiments by considering only values preceding the circles. 55

Figure 3.5 Prediction error (2.39) for the 0.5% and 0.75% bgc experiments obtained when using $\chi_\beta(t)$, $\chi_\beta^{\ln}(t)$ and $\chi_{NPP}(t)$ obtained from the 1% bgc experiment to predict the response. The error is shown as a function of the CO_2 final forcing strength. 58

Figure 3.6 Response function $\chi_\beta(t)$ derived by the three approaches (subfigure (a)) and the respective prediction errors for the first 30 years of the response (subfigure (b)). $\Delta c^{1\%}$: recovery with first approach from 1% bgc experiment; $\log(c)^{1\%}$: recovery with second approach from the 1% bgc experiment; $NPP^{1\%}_{polx}$: recovery with third approach from the 1% bgc experiment using for the derivative a polynomial fit of order x; $NPP^{2\times}_{polx}$: recovery with third approach from the $2\times CO_2$ bgc experiment using for the derivative a polynomial fit of order x. Continuous lines denote values of $\chi(t)$ that are within the time series length used for the recovery (30 years for $\Delta c^{1\%}$, 40 years for $\log(c)^{1\%}$ and 70 years for $NPP^{1\%}_{polx}$). Dotted lines denote extended parts of the response function, i.e. values not covered by the time series used for the recovery but obtained from the recovered spectrum using (2.12). Circles denote the response functions derived taking the full time series ($NPP^{2\times}_{polx}$). For more details see text. 61

Figure 3.7 Prediction of model responses employing response functions derived with the first approach from the 1% bgc experiment (derived from data with 30 years length but extended to 140 years) and with the third approach from the $2\times CO_2$ bgc experiment (derived from data with 140 years length). (a) Prediction by eq. (3.3) taking the response function $\chi_\beta(t)$ derived with the first approach. (b) Prediction by eq. (3.5) taking the response function $\chi_\beta(t)$ derived with the first approach and converted to $\chi_{NPP}(t)$ by eq. (3.9). (c) Prediction by eq. (3.5) taking the response function $\chi_{NPP}(t)$ derived in the first step of the third approach (see explanation after eq. (3.9)). Continuous lines are predictions and dashed lines are responses from the MPI-ESM. Circles indicate the maximum value for which responses are predictable according to our estimate of the linear regime (see text). The values printed to the right of the plots are the prediction errors (see eq. (2.39)) calculated for each experiment, considering when applicable only values preceding the circles. 62

Figure 3.8 Spectra associated to χ_{NPP} derived with different resolutions and Gregory plot for land carbon. (a) Spectrum derived with the first approach from the 1% bgc experiment; (b) Spectrum derived with the third approach from the $2\times CO_2$ bgc experiment; (c) "Gregory plot" for land carbon. Values of b indicate the rate at which the time derivative of land carbon changes with respect to the land carbon itself. Ranges of time scales corresponding to each rate accounting for one standard deviation are shown in subfigure (b). 65

Figure 3.9 Investigation of the land carbon response in the tropics and extra-tropics and how the regional response functions combine to the global response functions. The analysis is based on the $2\times CO_2$ bgc experiment. (a) Regional response functions; (b) Laplace transform $\widetilde{\chi}_{NPP}$ of global $\widetilde{\chi}_{NPP}$ obtained directly from the global carbon response (solid line) and from combining the tropical and extra-tropical response functions (dashed line); (c) As (b) but for χ_β; (d) As (c) but for q_β; (e) Tropical and extra-tropical spectra. For more details see text. 67

Figure 3.10 Investigation of differences in the response functions for tropics and extra-tropics. (a) Modified tropical spectrum by taking values of the extra-tropical spectrum from $\tau = 63$ years onwards; (b) resulting modified response function; (c) Modified tropical spectrum by taking values of the extra-tropical spectrum from $\tau = 63$ to $\tau = 251$ years; (d) resulting modified response function. First row shows that the difference in the response functions is associated to time scales greater than 63 years. The second row shows that the difference in the decay is related to time scales between 63 and 251 years, while time scales greater than 251 years change mostly the offset. 70

Figure 4.1 Generalized sensitivity $\chi_\beta^{(O)}(t)$ and prediction of responses from additional experiments. (a) Prediction error (4.15) for $\chi_\beta^{(O)}(t)$ derived with the first approach employing eq. (4.13) for prediction. (b) Prediction error (4.15) for $\chi_\beta^{(O)}(t)$ derived with the second approach employing eq. (4.14) for prediction. (c) Response function $\chi_\beta^{(O)}(t)$ recovered with the first approach at optimal forcing strength (Δc), second approach at optimal forcing strength ($\ln(c)$, opt. forcing strength), and second approach at maximal forcing strength ($\ln(c)$, max. forcing strength). (d) Prediction of additional experiments taking the "best" recovery of $\chi_\beta^{(O)}(t)$ (using the second approach at optimal forcing strength) employed in eq. (4.13). (e) Prediction of additional experiments taking the "best" recovery of $\chi_\beta^{(O)}(t)$ (using the second approach at optimal forcing strength) employed in eq. (4.14). Continuous lines are predictions and dashed lines are responses from the MPI-ESM. Dots indicate the maximum value for which responses are predictable by eq. (4.13) according to the estimate of the linear regime (see text). For better visibility of the regions within the linear regime in (c), the responses are shown only for the first 30 years. For more details see text. 84

Figure 4.2 Generalized sensitivity $\chi_\gamma^{(O)}(t)$ and prediction of responses from additional experiments. (a) Prediction error (4.15) employing the recovered $\chi_\gamma^{(O)}(t)$ in eq. (4.16). (b) Recovered response function $\chi_\gamma^{(O)}(t)$. (c) Prediction of additional experiments employing the recovered $\chi_\gamma^{(O)}(t)$ in eq. (4.16). Continuous lines are predictions and dashed lines are responses from the MPI-ESM. For more details see text. 88

Figure 4.3 Generalized sensitivity $\chi_\alpha^{(O)}(t)$ and prediction of responses from additional experiments. (a) Prediction error (4.15) for $\chi_\alpha^{(O)}(t)$ derived with the first approach employing eq. (4.5) for prediction. (b) Prediction error (4.15) for $\chi_\alpha^{(O)}(t)$ derived with the second approach employing eq. (4.18) for prediction. (c) Response function $\chi_\alpha^{(O)}(t)$ recovered with the second approach both enforcing and not enforcing monotonicity. (d) Prediction of additional experiments taking the "best" recovery of $\chi_\alpha^{(O)}(t)$ (using the second approach enforcing monotonicity) employed in eq. (4.5). (e) Prediction of additional experiments taking the "best" recovery of $\chi_\alpha^{(O)}(t)$ (using the second approach enforcing monotonicity) employed in eq. (4.18). Thick lines are predictions and thin lines are responses from the MPI-ESM. Dots indicate the maximum value for which responses are predictable by eq. (4.5) according to the estimate of the linear regime. For more details see text. 90

Figure 4.4 Generalized sensitivity $\chi_\alpha^{(L)}(t)$ and prediction of responses from additional experiments. (a) Prediction error (4.15) for $\chi_\alpha^{(L)}(t)$ derived with the first approach employing eq. (4.4) for prediction. (b) Prediction error (4.15) for $\chi_\alpha^{(L)}(t)$ derived with the second approach employing eq. (4.17) for prediction. (c) Response function $\chi_\alpha^{(L)}(t)$ recovered with the second approach both enforcing and not enforcing monotonicity. (d) Prediction of additional experiments taking the "best" recovery of $\chi_\alpha^{(L)}(t)$ (using the second approach enforcing monotonicity) employed in eq. (4.4). (e) Prediction of additional experiments taking the "best" recovery of $\chi_\alpha^{(L)}(t)$ (using the second approach enforcing monotonicity) employed in eq. (4.17). Thick lines are predictions and thin lines are responses from the MPI-ESM. Dots indicate the maximum value for which responses are predictable by eq. (4.4) according to the estimate of the linear regime. For more details see text. 91

Figure 4.5 Response function $\chi_\zeta(t)$ recovered from impulse-emission experiment. (a) Impulse response and fit by $\chi_\zeta(t)$. (b) $\chi_\zeta(t)$ recovered with and without enforcing the constraint $\chi_\zeta(0) = 1$. (c) Prediction of different experiments employing the recovered $\chi_\zeta(t)$ in eq. (4.19). For more details see text. 96

Figure 4.6 Quality of agreement between the airborne fraction determined from emission-driven simulations and that determined from the generalized α-β-γ framework. The close agreement shows that generalized α-β-γ framework gives a reasonable description of the linear dynamics of the coupled global carbon cycle in the MPI-ESM. 99

Figure 4.7 Generalized sensitivities in CMIP5 models. 101

Figure 4.8 Check of the additivity of the biogeochemical and radiative carbon responses in CMIP5 models. Plotted are the sum of the responses from the 1% bgc and 1% rad experiments (dashed lines) against the response from the 1% fully-coupled experiment (continuous lines) for land, ocean, and global carbon (land plus ocean). Additivity holds approximately within the estimated linear regime of 94 ppm for all models. For more details see text. 104

Figure 4.9 Airborne fraction and climate-carbon cycle feedbacks in CMIP5 models. The numbers in subfigure (b) indicate the model mean for each feedback function at 10-years and 100-years time scales. 106

Figure 4.10 Analysis of model spread and magnitude of feedback functions and their influence on the airborne fraction. (a) Standard deviation for each feedback function; (b) standard deviation for the airborne fraction and decomposition in terms of the feedback contributions according to eq. (4.26); (c) airborne fraction as presented in Fig. 4.9 (unchanged) and recalculated with the feedback function $\widetilde{f}_\beta^{(L)}$ for all models set to the model mean; (d) airborne fraction as presented in Fig. 4.9 (unchanged) and recalculated with the land biogeochemical feedback for all models set to zero, and with the ocean biogeochemical feedback for all models set to zero. See text for more details. 109

Figure A.1 $\chi(t)$ and $\mathbf{q}_\lambda$ derived for different values of M, $\log \tau_{min}$ and $\log \tau_{max}$ ($SNR \sim \mathcal{O}(10^5)$). As long as the discrete distribution of time-scales approximates densely enough the spectrum, the derived $\chi_\lambda(t)$ and $\mathbf{q}_\lambda$ are approximately independent of the values of the prescribed parameters. 129

Figure A.1 (continued) $\chi(t)$ and $\mathbf{q}_\lambda$ derived for different values of M, $\log \tau_{min}$ and $\log \tau_{max}$ ($SNR \sim \mathcal{O}(10^5)$). As long as the discrete distribution of time-scales approximates densely enough the spectrum, the derived $\chi_\lambda(t)$ and $\mathbf{q}_\lambda$ are approximately independent of the values of the prescribed parameters. 130

Figure A.1 (continued) $\chi(t)$ and $\mathbf{q}_\lambda$ derived for different values of M, $\log \tau_{min}$ and $\log \tau_{max}$ ($SNR \sim \mathcal{O}(10^5)$). As long as the discrete distribution of time-scales approximates densely enough the spectrum, the derived $\chi_\lambda(t)$ and $\mathbf{q}_\lambda$ are approximately independent of the values of the prescribed parameters. 131

Figure A.1 (continued) $\chi(t)$ and $\mathbf{q}_\lambda$ derived for different values of M, $\log \tau_{min}$ and $\log \tau_{max}$ ($SNR \sim \mathcal{O}(10^5)$). As long as the discrete distribution of time-scales approximates densely enough the spectrum, the derived $\chi_\lambda(t)$ and $\mathbf{q}_\lambda$ are approximately independent of the values of the prescribed parameters. 132

Figure B.1 Recovery of spectrum and $\chi(t)$ taking a true spectrum with contributions at time scales much longer and much shorter than the time scales covered by data. For the recovery, data with $SNR = 6 \times 10^4$ from the 1% experiment with the toy model (section 2.4.1) was taken (times series length T = 140 and time step Δt = 1). Parameters are taken as in section 2.4.2 except for $M = 140$. The derived spectrum recovers only partially the true spectrum. Nevertheless, the response function $\chi(t)$ is almost perfectly recovered. 141

Figure B.2 Spectrum $\mathbf{q}_\lambda$ and response function $\chi(t)$ derived for different discrete spectra $\mathbf{q}^*$ for $SNR \sim \mathcal{O}(10^4) - \mathcal{O}(10^5)$. Subfigures (a)-(f) show results for taking data from a 1% experiment and subfigures (g)-(j) for a $2 \times f_0$ experiment. The recovered spectra give smooth approximations to the underlying discrete spectra. The response functions are recovered almost perfectly. 144

Figure B.2 (continued) Spectrum $\mathbf{q}_\lambda$ and response function $\chi(t)$ derived for different discrete spectra $\mathbf{q}^*$ for $SNR \sim \mathcal{O}(10^4) - \mathcal{O}(10^5)$. Subfigures (a)-(f) show results for taking data from a 1% experiment and subfigures (g)-(j) for a $2 \times f_0$ experiment. The recovered spectra give smooth approximations to the underlying discrete spectra. The response functions are recovered almost perfectly. 145

Figure B.2 (continued) Spectrum $\mathbf{q}_\lambda$ and response function $\chi(t)$ derived for different discrete spectra $\mathbf{q}^*$ for $SNR \sim \mathcal{O}(10^4) - \mathcal{O}(10^5)$. Subfigures (a)-(f) show results for taking data from a 1% experiment and subfigures (g)-(j) for a $2 \times f_0$ experiment. The recovered spectra give smooth approximations to the underlying discrete spectra. The response functions are recovered almost perfectly. 146

Figure C.1 Recovery of $\chi_\gamma^{(O)}(t)$ by the RFI method and by the step method (Ragone et al., 2016). 149

Figure C.2 Decomposition of airborne fraction given by eq. (C.19) compared against eq. (4.21). 154

ACRONYMS

C^4MIP	Coupled Climate-Carbon Cycle Model Intercomparison Project
$CMIP_5$	Coupled Model Intercomparison Project Phase 5
ECS	Equilibrium Climate Sensitivity
ESM	Earth System Model
GPP	Gross Primary Production
MPI-ESM	Max Planck Institute for Meteorology Earth System Model
NPP	Net Primary Production
RFI	Response Function Identification
TCR	Transient Climate Response
TCRE	Transient Climate Response to Cumulative CO_2 Emissions

1

INTRODUCTION

This dissertation presents an attempt to systematically characterize the response of the global carbon cycle to atmospheric CO_2 perturbations in Earth System Models. The approach is to consider weak perturbations and describe the macroscopic response of the carbon cycle by means of linear response functions. In particular, I focus on the functions that quantify the so-called climate-carbon cycle feedbacks. By deriving and examining these functions, I show that one can obtain insight into the dynamics of the carbon cycle and its interaction with climate in these models.

The effect of atmospheric CO_2 on climate was first quantified by Arrhenius, 1896 already in the 19th century (Uppenbrink, 1996, Anderson et al., 2016). At that time, he reportedly estimated that it would take about 3,000 years for atmospheric CO_2 concentrations to double as a consequence of emissions from fossil fuel combustion (Uppenbrink, 1996). Today, barely a century later, we are almost halfway through: Atmospheric CO_2 levels increased by about 47% from pre-industrial times to 2018 (Friedlingstein et al., 2019). This increase is considered to be the dominant contribution to the radiative forcing driving climate change (IPCC, 2013). But changes in atmospheric CO_2 are not simply determined by anthropogenic emissions. They depend on the balance between emissions and changes in a vast natural cycle known as the global carbon cycle. Therefore, understanding the dynamics of this cycle, and in particular how it reacts as CO_2 is emitted, is critical in advancing climate research (Marotzke et al., 2017). Unfortunately, despite large scientific effort, many aspects of the global carbon cycle remain poorly understood (Ilyina and Friedlingstein, 2016).

The most advanced tools used to investigate the global carbon cycle and its interaction with climate are known as Earth System Models, which are complex mathematical models that simulate the coupling between these two systems. To try to better understand the behaviour of the carbon cycle in such complex models, a typical approach is to characterize the reaction of the global carbon as atmospheric CO_2 or emissions evolve according to a particular *scenario*, or a particular trajectory in time (Friedlingstein et al., 2006, Arora et al., 2013, Boer and Arora, 2013, Jones et al., 2013, Schwinger et al., 2014, Arora et al., 2019). This type of strategy, combined with observation-based studies, has led to a number of substantial advances in carbon-cycle research (Friedlingstein, 2015). Nevertheless, characterizing the behaviour of a system when forced in a particular way has the disadvantage that conclusions drawn from such analysis are not necessarily valid in situations where the forcing is different. For example, although observations show that the airborne fraction – the ratio of atmospheric CO_2 to anthropogenic emissions – has remained nearly constant for decades (Le Quéré et al., 2009, Raupach et al., 2014, Bennedsen et al., 2019), it has been argued

that this constancy holds only while emissions grow exponentially (Bacastow and Keeling, 1979, Raupach, 2013). This means that a constant airborne fraction is not a property of the carbon cycle, but a reflection of the response of the system when forced exponentially. Therefore, although the airborne fraction, quantified in this way, gives information on how the system is behaving under this specific perturbation, it does not necessarily inform about how the system may behave if emissions are for instance reduced, thus deviating from this exponential curve. A similar scenario dependence arises in the quantification of the climate-carbon cycle feedbacks that motivates the work in this dissertation (Gregory et al., 2009, Arora et al., 2013, Jones and Friedlingstein, 2020; see below). Hence, the proposal here is to try to characterize the behaviour of the carbon cycle in these models not in a particular scenario, but from a more general point of view. The idea is to identify characteristics that are, in a sense, *inherent to the system*, and not a result of a combination of the system and a particular scenario. Such approach permits a more systematic investigation of the dynamics of the carbon cycle, from which more general conclusions may be drawn. But before we get into the details of this analysis, let us explore a bit more the context within which this idea emerges.

1.1 THE GLOBAL CARBON CYCLE AND CLIMATE-CARBON CYCLE FEEDBACKS

The global carbon cycle can be understood as the flow of carbon through four large compartments of the Earth System: Atmosphere, land, ocean, and reserves in rocks and sediments (Carlson et al., 2001, Houghton, 2003, Ciais et al., 2013). Carbon in the atmosphere is encountered mostly in the form of CO_2, with smaller concentrations of methane, carbon monoxide and non-methane hydrocarbons. The atmosphere mixes within a relatively short time, so that CO_2 concentrations are approximately homogeneous around the planet (Houghton, 2003; Myhre et al., 2013, Box 8.2). From the point of view of the global carbon cycle, the atmosphere plays an important role in coupling the land and ocean compartments (Post et al., 1990).

Carbon on the land is stored essentially in vegetation in various forms, such as leaves, wood, stem, roots, and in dead organic matter in litter and soil (Ciais et al., 2013). On the land, as well as in the ocean, the carbon cycle is governed by a multitude of processes that develop over a wide range of time scales (Ciais et al., 2013, Box 6.1; Joos et al., 2013). In a simplistic view, the land carbon cycle can be understood as follows: Initially, carbon is taken up by plants through the process of photosynthesis, which, using water and light energy, converts CO_2 into carbohydrates and releases oxygen (Schlesinger, 1991, Singhal et al., 2012). Of this initial carbon uptake, known as gross primary production, a fraction is respired back to the atmosphere via autotrophic respiration. The difference between gross primary production and autotrophic respiration gives the net primary production, which is the net amount of carbon available for building plant tissues. Eventually, these tissues become dead organic matter, most of which is transferred to litter and soil compartments. The carbon is ultimately released to the atmosphere via heterotrophic respiration and disturbances such as natural and anthropogenic fires (Prentice et al., 2001, Ciais et al., 2013).

In the ocean, carbon is found mainly in the form of dissolved inorganic carbon, which consists of dissolved CO_2, carbonic acid, bicarbonate and carbonate ions (Mackenzie and Lerman, 2006, Ciais et al., 2013). In smaller quantities there is also dissolved organic carbon and marine biota, the majority of which being phytoplankton and other microorganisms (Ciais et al., 2013). The ocean-atmosphere carbon flux is driven by differences in partial pressure of CO_2 between the atmosphere and surface water (Field and Raupach, 2004, Prentice et al., 2001, Ciais et al., 2013). When taken up by the ocean, CO_2 dissolves and forms dissolved inorganic carbon, which is transported by currents and turbulent mixing (Heinze et al., 2015, Keppler, 2020). The ocean carbon dynamics is also affected by biological processes. For instance, phytoplankton assimilates CO_2 in the upper ocean via photosynthesis, thereby reducing the partial pressure of CO_2 at the surface and allowing for more CO_2 uptake from the atmosphere. On the other hand, the organic matter formed by photosynthesis might sink and be remineralized by microorganisms, releasing inorganic carbon to the interior ocean (Sarmiento and Bender, 1994, Heinze et al., 2015, Keppler, 2020).

The natural carbon cycle in the reserves in rocks and sediments is governed by extremely slow processes, such as weathering and sedimentation, that develop over time scales of the order of 10,000 years or more (Houghton, 2003, Ciais et al., 2013). Natural fluxes of carbon between this compartment and the three compartments described above are relatively small (Archer, 2010, Ciais et al., 2013), so that here these fluxes will not be considered. For our purposes, the relevance of this slow compartment lies in the fact that it contains a large reserve of fossil fuels. These fossil fuels have been systematically extracted and burned since the beginning of the industrial revolution, causing a major perturbation to atmospheric CO_2 and as a consequence to the land and ocean carbon cycle (Houghton, 2003, Ciais et al., 2013).

According to the Global Carbon Budget (Friedlingstein et al., 2019), in the period of 1850-2018 the cumulated amount of carbon emitted to the atmosphere was of 645 ± 65 PgC[1]. From this amount, about half was taken up by the land and ocean carbon compartments. Hence, the land and ocean carbon cycle have been playing a crucial role in avoiding the accumulation of even more carbon in the atmosphere, which could then lead to stronger climatic changes. This observation leads to important questions, such as how this uptake is taking place, and whether it will continue in a similar way in the future (Archer, 2010, Friedlingstein, 2015, Ilyina and Friedlingstein, 2016). To answer these questions, the understanding of the fundamental mechanisms driving the carbon cycle must be improved.

One way to proceed in this direction is by taking a macroscopic point of view and investigating the behaviour of the carbon cycle at a global scale. Insight into this behaviour can be gained by considering the concept of feedbacks. When a system is perturbed, a feedback can be seen as a response of the system that acts also as an additional perturbation, which can either amplify or reduce the initial perturbation. A simple example in the field of climate science is that of the ice-albedo feedback (Curry et al., 1995): Rising temperatures lead to ice melting; as ice melts, it reveals an underlying darker surface (e.g. soil); this

1 1 PgC = 10^{15} grams of carbon.

darker surface absorbs more radiation than ice, thereby amplifying the initial warming. In a similar way, feedback processes exist also in the carbon cycle. As emissions change atmospheric CO_2 concentrations, they cause a response in the land and ocean carbon cycle. This response may amplify or reduce the initial change in atmospheric CO_2. For example, when CO_2 concentrations increase, this increase leads to warming; with higher temperatures, heterotrophic respiration rates are enhanced (Raich and Potter, 1995). Thereby, more CO_2 is emitted, which amplifies the initial increase in CO_2. Because the initial increase is amplified, we call this process a positive feedback. Negative feedbacks also exist. For instance, as atmospheric CO_2 concentrations increase, the difference in partial pressure between CO_2 in the atmosphere and in the surface ocean increases, which drives CO_2 into the ocean. This in turn reduces the initial increase in atmospheric CO_2, and is therefore a negative feedback. Such feedbacks between climate and the carbon cycle will be here termed *climate-carbon cycle feedbacks*.

1.2 THE α-β-γ FRAMEWORK AND ITS GENERALIZATION

In principle, if one can quantify the response of the global carbon cycle in terms of these feedbacks, it is possible to gain insight into the relative contribution of the different processes associated with the feedbacks to the global dynamics of the system. A framework for such quantification was proposed by Friedlingstein et al., 2003. Essentially, the idea of Friedlingstein's framework is to quantify the response of global land and ocean carbon in terms of two global-scale feedbacks (Field and Raupach, 2004): A radiative feedback, which collectively quantifies feedbacks arising from processes that respond to changes in climate (as in the example of heterotrophic respiration); and a biogeochemical feedback, which collectively quantifies feedbacks arising from processes that respond to changes in CO_2 concentration (as in the example of ocean uptake). In addition to the feedbacks, one also accounts for the climate response to changes in CO_2. These two feedbacks and the climate response are quantified in this framework by three *sensitivities*, respectively γ, β, and α. For this reason, this framework is also known as the α-β-γ framework.

The α-β-γ framework led to important insights into the role of feedbacks in the carbon cycle and stimulated a vast amount of research in this field (Friedlingstein et al., 2006, Gregory et al., 2009, Arneth et al., 2010, Zickfeld et al., 2011, Boer and Arora, 2013, Arora et al., 2013, Schwinger et al., 2014, Friedlingstein et al., 2014, Friedlingstein, 2015, Adloff et al., 2018, Williams et al., 2019, Goodwin et al., 2019, Jones and Friedlingstein, 2020). Nevertheless, significant limitations in the framework were identified (Gregory et al., 2009, Arora et al., 2013, Boer and Arora, 2013). One serious limitation is that the α-β-γ sensitivities and therefore the whole framework are scenario dependent, meaning that the quantification of the feedbacks is in principle only valid for the scenario that is considered for the calculations. This leads to a situation similar to that described for the airborne fraction at the beginning of this chapter: Although these feedbacks inform about the behaviour of the system under a particular scenario, they do not give information about

how the system may behave if one considers a different scenario. This therefore hinders a more general understanding of the dynamics of these feedbacks.

Hence, the question arises whether it is possible to systematically characterize the α-β-γ sensitivities, and thereby the feedbacks, so that the results are valid for all perturbation scenarios. It turns out that by considering weak perturbations, such a generalization is possible (Rubino et al., 2016, Enting and Clisby, 2019). The idea behind the generalization is the following: As recognized by Gregory et al., 2009, Friedlingstein's framework assumes an instantaneous response of the land and ocean carbon cycle to perturbations. Nevertheless, it is well-known that the processes that govern the dynamics of carbon in these compartments develop over a wide range of time scales. This essentially means that the system has a "memory", i.e. that present values of the response of the system depend not only on the present values of the perturbation, but also on its past values. The generalization proposed by Rubino et al., 2016 and Enting and Clisby, 2019 accounts for this memory. The result is that the sensitivities are not any more numbers, but functions; more precisely, *linear response functions*. The resulting feedbacks are characterized in terms of their time-scale dependence and are independent of the scenario. In addition, from the feedbacks it is possible to obtain a time-scale dependent form of the airborne fraction that informs about the behaviour of atmospheric CO_2 under all emission scenarios. This framework therefore provides a formalism to systematically investigate the climate-carbon cycle feedbacks and their influence on the airborne fraction.

1.3 MOTIVATING RESEARCH QUESTIONS

Although the referred authors have proposed this generalized α-β-γ framework, no attempt has been made so far to systematically investigate its applicability to Earth System Models. It is the question of this applicability that constitutes the main motivation for the present work.

As will be seen, this question poses serious challenges, and requires the development of a specific method that allows for the practical application of this framework. But although this method is developed with this motivation in mind, its relevance is not restricted to the problem of the climate-carbon cycle feedbacks. This is because the generalization of the α-β-γ sensitivities can be seen as a particular application of the linear response framework (Kubo, 1957), which is a general approach to characterize the linear response of a system. This approach has been rigorously investigated from a mathematical point of view (Ruelle, 2009) and has applications in several fields (e.g. Zwanzig, 2001, Lucarini and Sarno, 2010, Kubo et al., 2012, Ragone et al., 2016, Lucarini et al., 2017, Aengenheyster et al., 2018, Lembo et al., 2020, Bódai et al., 2020, Gottwald, 2020). Insofar that the method developed here contributes to the practical application of the linear response framework, this method may be helpful not only to investigate the questions proposed in this dissertation but also to study different problems where this framework can be applied. The relationship between the linear response framework and the generalization of the α-β-γ sensitivities, as well as

some potential applications of the method developed here to problems in climate science are highlighted in the introductory sections of chapter 2.

The work in this dissertation proceeds in three steps. The first consists in the development of the method. As will be discussed, the main difficulty in applying the generalized α-β-γ framework lies in the identification of the linear response functions that generalize the α-β-γ sensitivities. Since the goal is to investigate the applicability of the generalized framework to Earth System Models, one must be able to derive these functions taking published data from standard C^{4}MIP (Coupled Climate-Carbon Cycle Model Intercomparison Project) perturbation experiments. Nevertheless, methods available in the literature to derive response functions from data require tailored perturbation experiments, and are therefore not applicable to C^{4}MIP data. Therefore, chapter 2 deals with the general problem of how to derive linear response functions taking data from arbitrary perturbation experiments. In this chapter I address the following research question:

Q2: How to identify linear response functions from arbitrary perturbation experiments?

If one hopes to apply the generalized α-β-γ framework to Earth System Models, a method to derive the generalized sensitivities must be demonstrated to work in application to real data from C^{4}MIP experiments performed with an Earth System Model. This is the main purpose of chapter 3. In this chapter I employ the method developed in chapter 2 to identify the linear response functions that generalize the β and γ sensitivities for the land carbon in the Max Planck Institute for Meteorology Earth System Model taking data from standard 1% experiments. Further, I estimate the linear regime for which the generalization is valid. The robustness of the identified generalized sensitivities is demonstrated by their ability to predict the response of the model in several experiments. Moreover, I show another application of the method developed in chapter 2. I demonstrate how by this method one can obtain information on the internal dynamics of the carbon cycle by recovering the spectrum of time scales of the response. I test the robustness of the recovered spectrum by two independent methods, and illustrate how one can investigate the reasons behind the structure of the spectrum. Thus this chapter is motivated by the following questions:

Q3.1: Can one robustly identify from C^{4}MIP-type experiments the linear response functions that generalize the land carbon sensitivities in the MPI-ESM? What is the range of linearity for which the generalization is valid?

Q3.2: Is it possible to obtain information about the internal dynamics of the carbon cycle from these functions?

In chapter 4, I investigate the generalized α-β-γ framework in Earth System Models. This chapter has two main purposes. The first is to demonstrate that the generalized α-β-γ framework correctly describes the linear dynamics of the global carbon cycle in an Earth System Model. To this end, I compare the carbon cycle dynamics in the MPI-ESM as described by the framework with the behaviour of the model as observed from simulations.

The second purpose of this chapter is to quantify the feedbacks and the airborne fraction in different Earth System Models. For this purpose I employ the framework to calculate these quantitites for an ensemble of CMIP5 (Coupled Model Intercomparison Project Phase 5) models. In this study it is investigated how the feedbacks determine the dynamics of the airborne fraction at different time scales, and how the model spread in the feedbacks influences the spread in the airborne fraction. Therefore in this chapter the guiding questions are:

Q4.1: Can the generalized α-β-γ framework describe with reasonable accuracy the linear dynamics of the global carbon cycle in an Earth System Model?

Q4.2: How are the feedbacks and airborne fraction determined at different time scales in different Earth System Models? What can we conclude from them?

In chapter 5 I summarize my findings, provide a critical assessment of the results, and discuss their implications within the broader scientific context.

2

IDENTIFICATION OF LINEAR RESPONSE FUNCTIONS FROM ARBITRARY PERTURBATION EXPERIMENTS

2.1 INTRODUCTION

To better understand the response of climate and the carbon cycle to perturbations from atmospheric CO_2 (e.g. due to fossil fuels combustion and land use change), internationally experiments with Earth System Models have been performed under a common protocol within the Coupled Climate Carbon Cycle Model Intercomparison Project (C^4MIP; Fung et al., 2000, Friedlingstein et al., 2006) that today is part of the International Coupled Model Intercomparison Project (CMIP; Taylor et al., 2012). While the response of ocean carbon agrees well among models, the land carbon response shows a large spread (Friedlingstein et al., 2006, Arora et al., 2013). The spread comes from differences in the implementation of processes (such as carbon assimilation, carbon allocation, heterotrophic respiration, water and nutrients limitation, land cover change, fire, etc.) in different models. But trying to explain these differences by looking into implementation details of models is an extremely difficult task due to the complexity of Earth System Models. An alternative is to try to clarify the differences by comparing characteristics of the responses. In the case of land carbon, the response to CO_2 perturbations is typically characterized by the land carbon sensitivities γ and β, proposed by Friedlingstein et al., 2003. Considering land carbon (and not ocean carbon), γ quantifies the sensitivity of carbon to the effect of CO_2 on climate change, and β quantifies the sensitivity of carbon to the effect of CO_2 on photosynthetic productivity.

Nevertheless, the proposed framework of γ and β characterizes only the sensitivity for a particular perturbation scenario (Gregory et al., 2009, Arora et al., 2013). Instead, one would like to characterize the sensitivity as such. The dependence of the γ- and β-sensitivities on the considered scenario arises because of the memory of the system, i.e. the dependence of the response on past values of the perturbation (see Rubino et al., 2016 and Enting and Clisby, 2019). When the memory is taken into account, linear response functions arise as natural generalization of the climate-carbon cycle sensitivities that characterize the behaviour of the model to any type of perturbation (see below).

To follow this idea, the main challenge is to obtain the appropriate linear response functions. Although there are methods available in the literature, they require tailored perturbation experiments and are therefore not applicable to published C^4MIP data. Hence, the aim of the present study is to develop a method to identify linear response functions using data from any type of perturbation experiment. To demonstrate its applicability to

C^4MIP data, in chapter 3 this method will be employed to identify for the Max Planck Institute for Meteorology Earth System Model (MPI-ESM; see Appendix A.1) from standard 1% experiments the land carbon *generalized sensitivities*, i.e. the linear response functions that generalize the land carbon sensitivities γ and β.

In addition to climate-carbon cycle sensitivities, the method developed here may find application in a range of different problems. Ragone et al., 2016 have shown how the linear response framework generalizes the concept of equilibrium climate sensitivity (ECS) and transient climate response (TCR). Originally, ECS is defined as the response of global temperature to an instantaneous doubling of atmospheric CO_2 from its pre-industrial value, while TCR is the temperature response to a doubling of CO_2 after a 1% per year increase. Within the generalized framework, the ECS and TCR are shown to be only particular values encoded in the linear response function.

The linear response framework can be applied to predict the response of complex models whose simulations are computationally expensive. A recent study (Lembo et al., 2020) has shown the applicability of the approach to predict the response of the MPI-ESM in several climate variables such as globally averaged near-surface temperature, Atlantic meridional overturning circulation, Antarctic circumpolar current and ocean heat uptake. For the study, an ensemble of tailored perturbation experiments was employed. With the present method, one could perform such predictions taking only data from a given perturbation experiment and a control simulation.

Linear response functions can also be useful in studying "committed changes" (Wigley, 2005, Plattner et al., 2008, Mauritsen and Pincus, 2017; for a study focused on commited terrestrial ecosystem changes, see Jones et al., 2009). As shown by Ragone et al., 2016, the concept of climate inertia (closely related to commited changes) can be explicitly described within the linear response function framework. Since the linear response function describes the delayed response of the system to a perturbation, by deriving the linear response function one has at hand also the information of how the system reacts once the perturbation – for instance CO_2 emissions – ceases.

Additionally, linear response functions can help understanding the concept of "emergent constraints" (Nijsse and Dijkstra, 2018). Recent studies have shown how to obtain from response functions derived from conceptual models emergent constraints for the real Earth System (Cox et al., 2018, Williamson et al., 2019). With the method develop here, in principle such type of analyses may be carried out employing instead response functions derived from Earth System Models.

The theoretical background behind the linear response framework is given by linear response theory (Kubo, 1957). Although the existence of linear response functions has only been proven for hyperbolic systems (Ruelle, 1997, Ruelle, 1998a), the "chaotic hypothesis" (Gallavotti and Cohen, 1995) suggests that results for this class of systems transfer to much wider classes including chaotic high-dimensional systems such as the Earth system. While the application of linear response theory following the developments of Ruelle is a rather recent topic in climate science (e.g. Lucarini, 2009, Lucarini and Sarno, 2010, Lucarini et al.,

2014, Ragone et al., 2016, Lucarini et al., 2017, Aengenheyster et al., 2018, Ghil and Lucarini, 2019, Lembo et al., 2020, Bódai et al., 2020), linear response functions have been used to study climate already for decades (e.g. Siegenthaler and Oeschger, 1978, Emanuel et al., 1981, Maier-Reimer and Hasselmann, 1987, Enting, 1990, Joos and Bruno, 1996, Joos et al., 1996, Thompson and Randerson, 1999, Pongratz et al., 2011, Joos et al., 2013, Enting and Clisby, 2019). Regardless of how one chooses to view the problem – either through the lens of linear response theory or another type of linearization –, a fundamental step in the application of this approach is the identification of the linear response function.

In the field of climate science, the linear response function is typically identified by means of the impulse response function, which is the response to a Dirac delta-type perturbation (e.g. Siegenthaler and Oeschger, 1978, Maier-Reimer and Hasselmann, 1987, Joos et al., 1996, Thompson and Randerson, 1999, Joos et al., 2013). This method has become so widely known that often the terms linear response function and impulse response function are used interchangeably. Indeed, in the particular case where perturbations are weak, the two concepts coincide. But this is not true in general: If the impulse strength is large so that nonlinearities become important, the impulse response function differs from the linear response function.

Other studies have proposed to identify the linear response function by making use of other types of perturbations. Reick, 2002 and Lucarini, 2009 used a weak periodic forcing to derive response functions in the Fourier space (also called susceptibility). Hasselmann et al., 1993, Ragone et al., 2016, MacMartin and Kravitz, 2016, Lucarini et al., 2017, Van Zalinge et al., 2017, Aengenheyster et al., 2018 and Bódai et al., 2020 identify the linear response function using step experiments, where the perturbation is a Heaviside-type function.

As noted by Lucarini et al., 2014, in principle the linear response function of a system can be derived from an arbitrary type of perturbation experiment. A linear response function $\chi(t)$ is implicitly defined by a convolution equation of the type

$$R(t) = \int_0^t \chi(t-s)\Delta f(s)ds, \tag{2.1}$$

where R is the linear response to a forcing Δf. One method would be to apply a Laplace transform to (2.1), so that $\chi(t)$ can in principle be computed by the inverse Laplace transform

$$\chi(t) = \mathcal{L}^{-1}\{\mathcal{L}\{R\}/\mathcal{L}\{\Delta f\}\}, \tag{2.2}$$

where $\mathcal{L}\{\cdot\}$ is the Laplace transform operator. In fact, a first step towards the derivation of $\chi(t)$ from the general equation (2.1) was taken by Pongratz et al., 2011, although the problem was not systematically discussed.

Deriving $\chi(t)$ from perturbation experiments data is not a trivial problem. For the general case where the perturbation is different from a Dirac delta-type function, the problem is ill-posed. This basically means that attempts to recover the exact $\chi(t)$ yield a solution with large errors due to the noise in the data. On the other hand, when Δf is a Dirac delta-type

function with sufficiently small perturbation strength so that the response can be considered linear, the impulse response gives directly the linear response function, i.e. $\chi(t) = R(t)$. But even in this case noise may hinder the recovery: Because the perturbation is only one "pulse" with small perturbation strength, the response may have a too low signal-to-noise ratio, giving once more a recovery with large errors.

To remedy these noise problems, a method intended to "damp" the noise in the response is usually employed. In MacMartin and Kravitz, 2016, a step experiment with large perturbation strength is used to obtain a better signal-to-noise ratio in the response, but at the cost of enhancing the effect of nonlinearities. An alternative approach is employed by Ragone et al., 2016 and Lucarini et al., 2017, who employ an ensemble of simulations and take the ensemble averaged response so that the level of noise is reduced. But especially for complex models such as Earth System Models, ensembles of simulations can be extremely expensive so that such a procedure may not be feasible.

Instead of trying to improve the signal-to-noise ratio of the data, here the interest lies in deriving $\chi(t)$ from a *single* realization of a *given* experiment by *accounting for the ill-posedness* of the problem. For this purpose, I employ regularization theory. Although this theory offers a variety of methods to solve ill-posed problems (see e.g. Groetsch, 1984, Bertero, 1989, Bertero et al., 1995, Engl et al., 1996, Hansen, 2010), currently no general all purpose method exists. Typically, methods rely on some type of prior information about the problem (Istratov and Vyvenko, 1999). Hence, they must be tailored according to the particularities in each application. Here, I develop a method requiring only data from a given perturbation experiment and a control simulation (see section 2.3). The method solves the ill-posed problem by applying Tikhonov-Phillips regularization. The regularization parameter is chosen via the discrepancy method. An essential ingredient of the discrepancy method is the noise level, which is usually not known a priori. For this reason, I propose a method to estimate the noise level by taking advantage of the information given by a spectral analysis of the perturbation experiment and the control simulation. If the desired response function is known to be monotonic, the noise estimate can be further adjusted. In section 2.4, the method is demonstrated to give reliable results under appropriate conditions of noise and nonlinearity. In section 2.5, I compare the derived method with two existent methods in the literature to identify the response function in the time domain. Results and technical details are discussed in section 2.6. Additional calculations are shifted to appendices.

2.2 LINEAR RESPONSE THEORY AND GENERALIZATION OF CLIMATE-CARBON CYCLE SENSITIVITIES

The purpose of this study is to develop a method to derive linear response functions from C^4MIP-type experiments. These experiments consist of a long control experiment and in particular experiments where the CO_2 concentration is increased by a certain percentage each year over a span of 140 years (see Table 2.1 and Fig. 2.1 for the C^4MIP-type experi-

ments considered in this study). The idea behind these experiments is to characterize in a standardized way the sensitivity of the global carbon cycle of different Earth system models to CO_2 perturbations. Friedlingstein et al., 2003 introduced for this purpose the α-β-γ sensitivities. But α-β-γ characterize the carbon cycle sensitivity only for this particular type of C^4MIP-experiments. Since linear response functions can be understood as generalizations of these sensitivities that characterize the response to any type of perturbation, one would be interested to have a method to derive these functions as well from C^4MIP-type experiments, since data from these experiments are readily available in international databases.

This section prepares for the introduction of the method developed in this study by 1) demonstrating how linear response functions generalize the α-β-γ sensitivities (see also Rubino et al., 2016 and Enting and Clisby, 2019); and 2) introducing the basic ansatz underlying the method, which will be then derived in section 2.3.

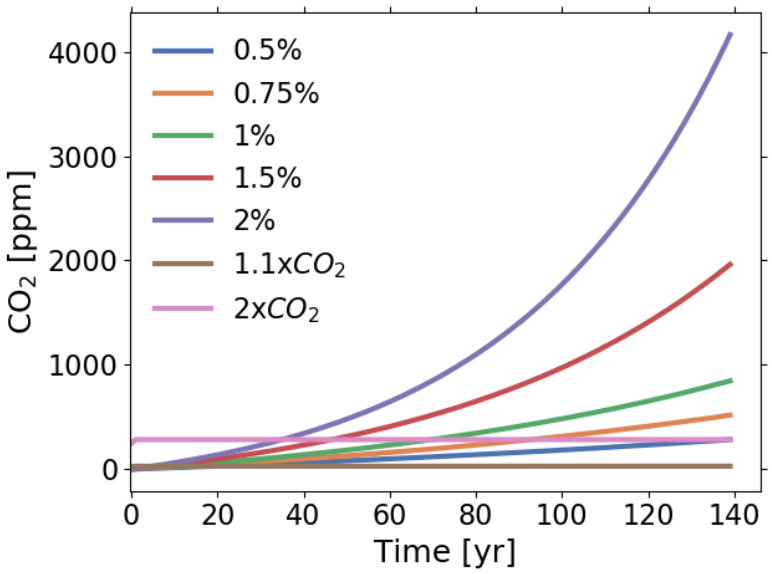


Figure 2.1 : Forcings for the CMIP-type experiments considered in this study.

That linear response functions generalize the Friedlingstein α-β-γ sensitivities can be seen as follows. I explain this by taking the β-sensitivity of land carbon uptake as an example. The calculation of β is based on an experiment where the CO_2-rise acts only biogeochemically, i.e. concerning the land carbon via the photosynthesis of plants. Calling $\Delta C^{bgc}(t)$ the difference in land carbon between the perturbed and the control simulation, β is defined as

$$\beta(t) := \frac{\Delta C^{bgc}(t)}{\Delta c(t)}, \quad \Delta c(t) = c_{PI}(1.01^t - 1), \tag{2.3}$$

where t is the time in years elapsed since the perturbation was switched on, and $\Delta c(t)$ is the change in CO_2 concentration when increasing atmospheric CO_2 each year by 1%

Table 2.1 : C^4MIP-type experiments considered in this study. Forcings are shown in Fig. 2.1. Acronyms "rad" and "bgc" refer to standard CMIP model setups used to calculate the climate-carbon cycle sensitivities. In the "rad" (radiatively coupled) setup only the radiation code of the model is affected by changes in atmospheric CO_2. This setup is used to calculate γ. In the "bgc" (biogeochemically coupled) setup only the biogeochemistry of the model is affected by changes in atmospheric CO_2. This setup is used to calculate β. In brackets are names of standard CMIP experiments. Note that because the toy model (section 2.4) has a purely numerical purpose such "rad" and "bgc" setups do not apply there.

Type	Forcing	Description
Percent	0.5% rad	CO_2 is increased from its pre-industrial value at the specified percent rate per year.
	0.5% bgc	
	0.75% rad	
	0.75% bgc	
	1% rad (esmFdbk1)	
	1% bgc (esmFixClim1)	
	1.5% rad	
	1.5% bgc	
	2% rad	
	2% bgc	
Step	1.1 $\times$ CO_2 rad	CO_2 is abruptly increased from its pre-industrial value by the specified factor.
	1.1 $\times$ CO_2 bgc	
	2 $\times$ CO_2 rad	
	2 $\times$ CO_2 bgc	
Control	pre-industrial (piControl)	CO_2 is held fixed at its pre-industrial value.

starting from its pre-industrial value c_{PI} of the control simulation. In the framework of linear response, one can understand $\Delta c(t)$ as perturbation and $\Delta C^{bgc}(t)$ as response so that the response formula (2.1) specialized to the present case reads

$$\Delta C^{bgc}(t) = \int_0^t \chi_\beta(t-s)\Delta c(s)ds. \tag{2.4}$$

This equation defines $\chi_\beta(t)$ as the linear response function describing the biogeochemical response of land carbon to any type of CO_2 perturbation $\Delta c(t)$, as long as the perturbation is sufficiently weak. Employing in particular the percent perturbation from (2.3), $\beta(t)$ is thus obtained from the linear response function $\chi_\beta(t)$ as

$$\beta(t) = \frac{1}{\Delta c(t)} \int_0^t \chi_\beta(t-s)\Delta c(s)ds. \tag{2.5}$$

In this way, $\chi_\beta(t)$ can be understood as a generalization of $\beta(t)$, that gives not only the response to percent-type perturbations, but also to other perturbations. Accordingly, $\chi_\beta(t)$ characterizes system properties independent of the type of perturbation, in contrast to $\beta(t)$.

So far, I have tacitly assumed the existence of appropriate linear response functions. But there is no reason that for general systems linear response functions should exist. Ruelle rigorously demonstrated the existence of linear response functions for perturbed hyperbolic systems. Even though hyperbolic systems are rather special, also non-hyperbolic systems may show a linear response (Reick, 2002, Lucarini, 2009) and it is believed that results for this class of systems transfer to large classes of nonequilibrium systems, including the climate system (Ruelle, 1999, Lucarini, 2008, Lucarini and Sarno, 2010, Gallavotti, 2014, Ragone et al., 2016, Lucarini et al., 2017). Ruelle considered systems of type

$$\frac{d}{dt}\mathbf{x} = \mathbf{A}_0(\mathbf{x}) + \mathbf{A}_1(\mathbf{x})\Delta f(t), \tag{2.6}$$

where $\mathbf{x}(t)$ is the possibly infinite dimensional state vector and the perturbation $\Delta f(t)$ couples to the unperturbed system $\frac{d}{dt}\mathbf{x} = \mathbf{A}_0(\mathbf{x})$ via the field $\mathbf{A}_1(\mathbf{x})$. In the present context eq. (2.6) could e.g. represent the dynamics of the Earth system perturbed by anthropogenic emissions $\Delta f(t)$. Considering an observable $Y(\mathbf{x})$ – which in our example is land carbon – Ruelle proved that the ensemble average of its deviation from the unperturbed system $\langle \Delta Y \rangle$ can be expanded in the perturbation $\Delta f(t)$:

$$\langle \Delta Y(\mathbf{x}(t)) \rangle = \int_0^t \chi(t-s)\Delta f(s)ds + \mathcal{O}((\Delta f)^2). \tag{2.7}$$

Hence, mathematically, in this framing linear response functions do not describe responses in the observables directly, but only in their ensemble averages. For the recovery of linear response functions from numerical experiments, this would mean that one had to perform many simulations to obtain the appropriate ensemble averages. Using tailored perturbations

experiments, it was demonstrated in several studies (e.g. Ragone et al., 2016, Lucarini et al., 2017, Bódai et al., 2020) that linear response functions can indeed be obtained in this way, but at the expense of a large numerical burden from the need to perform many experiments. Instead, the aim here is to obtain the linear response functions from a *given* experiment and only from a *single* realization. Since we are dealing with a single realization, (2.7) becomes

$$\Delta Y(t) = \int_0^t \chi(t-s)\Delta f(s)ds + \eta(t) + \mathcal{O}((\Delta f)^2), \tag{2.8}$$

where $\eta(t)$ is a noise term that must show up as a consequence of dropping the ensemble average. In addition, I assume linearity in the perturbation. As a consequence, the present study is based on the ansatz

$$\Delta Y(t) = \int_0^t \chi(t-s)\Delta f(s)ds + \eta(t), \tag{2.9}$$

where now the response $\Delta Y(t)$ is divided into a deterministic term $\int_0^t \chi(t-s)\Delta f(s)ds$ and a noise term $\eta(t)$.

The linearity assumption is by purpose: In the present approach to derive the linear response function (see next section), hereafter called *RFI method* (Response Function Identification method), I first use (2.9) to obtain $\chi(t)$ and justify the linearity assumption a posteriori by analyzing how robustly the response can be recovered for different perturbation strengths. Dropping the nonlinear terms has the advantage that one can use the corpus of linear methods to derive $\chi(t)$ from (2.9). Note that in practice, however small the perturbation may be, the nonlinear terms do not vanish. Therefore, the contribution of nonlinearities is in this way distributed among $\chi(t)$ and $\eta(t)$, which will be different from the previous $\chi(t)$ and $\eta(t)$ in (2.8). How strongly nonlinearities affect the numerical identification of $\chi(t)$ depends on the estimation of $\eta(t)$, which is a crucial part of our RFI method and the main novelty introduced here to deal with the ill-posedness of the problem to identify $\chi(t)$.

2.3 IDENTIFICATION OF LINEAR RESPONSE FUNCTIONS FROM ARBITRARY PERTURBATION EXPERIMENTS

In this section I derive the RFI method. As mentioned above, the aim of this method is to obtain the linear response function using data from a single realization of a given perturbation experiment. For this purpose, the estimation of the noise term $\eta(t)$ is an essential step, which requires additionally data from a control experiment.

Starting from the ansatz (2.9), the method is based on the idea that the noise term $\eta(t)$ can be estimated using information on the internal variability from the control experiment in combination with a spectral analysis of the perturbation experiment. The identification of the linear response function proceeds as follows: First, I assume that the linear response

function decays multi-exponentially, i.e. I take a functional form for $\chi(t)$. Second, (2.9) is discretized for application to the discrete set of data, which results in a matrix equation. Then, assuming that the solution obeys the Picard condition (see below), I estimate the high-frequency components of the noise term $\eta(t)$ in (2.9) via a spectral analysis of the matrix equation applied to the data from the perturbation experiment. Next, assuming that the spectral distribution of noise is similar for the control and perturbation experiments, I estimate also the low-frequency components $\eta(t)$. The final estimate of $\eta(t)$ is then used in a regularization procedure to find an approximate solution for $\chi(t)$. In case $\chi(t)$ is known to be monotonic, the approximated solution is further adjusted by checking for monotonicity.

In the first subsection, I introduce the assumption for the functional form of the linear response function. In subsections 2.3.2 and 2.3.3, I present the discretized problem and the solution using a regularization technique. In subsection 2.3.4 I derive the method to estimate the regularization parameter by estimating the noise in the data. Finally, in subsection 2.3.5 I show how the solution can be further improved in the presence of a monotonicity constraint.

2.3.1 *Functional form of the linear response function*

In general, the identification of linear response functions from data may be performed either pointwise (e.g. Ragone et al., 2016) or assuming a functional form (e.g. Maier-Reimer and Hasselmann, 1987). Both approaches usually lead to an ill-posed problem, and therefore to similar difficulties to find the solution (see more details in subsection 2.3.3). Although the RFI method may be applied in either case, here I assume that the response function consists of an overlay of exponential modes. By this ansatz I guarantee from the outset that the response relaxes to zero for $t \to \infty$, which is consistent with results from simulation of the global carbon cycle (see e.g. Joos et al., 2013). Besides constraining the function space for the derivation of the response function, this gives in principle also the spectrum of internal time scales of the response.

Assuming this ansatz, the question on the functional form of $\chi(t)$ arises. In climate science, it is typically assumed that the response function can be described by only a few exponents (Maier-Reimer and Hasselmann, 1987, Enting and Mansbridge, 1987, Hasselmann et al., 1993, Hasselmann et al., 1997, Grieser and Schönwiese, 2001, Li and Jarvis, 2009, Joos et al., 2013, Colbourn et al., 2015, Lord et al., 2016), i.e.

$$\chi(t) := \sum_{i=1}^{M} g_i e^{-t/\tau_i} \quad \text{for } M \text{ small,} \tag{2.10}$$

where the τ_i values are interpreted as characteristic time scales and the g_i values are their respective weights. τ_i and g_i are then obtained by applying some fitting technique taking a fixed number of terms M. Thus, an important step in this type of approach is to determine

a suitable value for M. A common practice is to prescribe M starting from a small number, solve the problem and then add terms progressively, until the addition of a new term does not anymore improve the fit according to some quality-of-fit criterion (e.g. Kumaresan et al., 1984, Maier-Reimer and Hasselmann, 1987, Hasselmann et al., 1993, Pongratz et al., 2011, Colbourn et al., 2015, Lord et al., 2016). Thereby it is assumed that the information in the data has been already fully exploited so that fitting of additional terms would be artificial. Nevertheless, finding the parameters τ_i and g_i either from a given $\chi(t)$ by (2.10) or from $\Delta Y(t)$ by inserting (2.10) into (2.1) means to solve a Fredholm equation of the first kind, which is an ill-posed problem (Groetsch, 1984, Istratov and Vyvenko, 1999). This implies that even though the obtained solution may give a very good fit to the data, it may significantly differ from the exact solution (see e.g. the famous example from Lanczos, 1956).

Therefore, to avoid the complication of determining M, I assume instead that the response function is characterized by a continuous spectrum $g(\tau)$ (Forney and Rothman, 2012a):

$$\chi(t) = \int_0^\infty g(\tau) e^{-t/\tau} d\tau. \tag{2.11}$$

By this approach the time scale τ is not anymore an unknown, but is given after discretization by a prescribed distribution with M terms covering a wide range of τ_i values. Thus, only a discrete approximation to the spectrum $g(\tau)$ needs to be found. In this way, although the distribution of time scales is prescribed, the obtained discrete spectrum gives a good approximation to the underlying continuous spectrum independently of the number M, provided that the extent and resolution of the prescribed distribution approximate the underlying range of time scales sufficiently densely and that the quality of the data is sufficiently good (see a numerical demonstration in Appendix A.2).

This approach has two additional advantages. First, by prescribing the distribution of time scales one must solve instead of a nonlinear ill-posed problem – as when obtaining τ_i and g_i from (2.10) – only a linear ill-posed problem, for which the mathematical theory is fairly well developed (Groetsch, 1984, Engl et al., 1996). Because the problem is linear, the solution is given analytically (see section 2.3.3), which makes the method more transparent. Second, because in the application to the carbon cycle intended here I will deal with variables integrated over many climate zones, one expects to obtain not few time scales but rather a large spectrum of time scales, which is obviously better represented by (2.11) than by (2.10).

2.3.2 *Discretized problem*

Due to the wide range of time-scales in the carbon cycle (Ciais et al., 2013, Box 6.1), it is useful to switch to a logarithmic scale (Forney and Rothman, 2012a) by rewriting (2.11) in terms of $\log_{10}\tau$:

$$\chi(t) = \int_{-\infty}^\infty q(\tau) e^{-t/\tau} d\log_{10}\tau, \quad \text{with} \quad q(\tau) := \tau \ln(10) g(\tau). \tag{2.12}$$

Hereafter, $q(\tau)$ and its discrete version $\mathbf{q}$ (see below) will be called *spectrum*.

In order to apply the basic equation (2.9) together with (2.12) to simulation data, the whole problem needs to be discretized in time and also with respect to the spectrum of time scales. Here I assume the data to be given at equally spaced time steps $t_k = t_0 + k\Delta t$, $k = 0, 1, \ldots, N-1$, where N is the number of data, while the time scales are assumed to be equally spaced at a logarithmic scale between maximum and minimum values τ_{max} and τ_{min}, i.e.

$$\begin{aligned} \log_{10}\tau_j = \log_{10}\tau_{min} + j\Delta\log_{10}\tau, \quad j = 0, 1, \ldots, M-1, \\ \text{with} \quad \Delta\log_{10}\tau := \frac{\log_{10}\tau_{max} - \log_{10}\tau_{min}}{M} \end{aligned} \tag{2.13}$$

where M is the number of time scales. As shown in Appendix A.3, the resulting discretized equations corresponding to (2.9) and (2.12) are

$$\Delta Y_k = \Delta t \sum_{i=0}^{k} \chi_{k-i}\, \Delta f_i + \eta_k, \quad k = 0, ..., N-1, \tag{2.14}$$

and

$$\chi_k = \Delta\log_{10}\tau \sum_{j=0}^{M-1} q_j e^{-k\Delta t/\tau_j}, \quad k = 0, ..., N-1, \tag{2.15}$$

where η_k stands for the noise. Combining the response data ΔY_k, the spectral values q_j, and the noise values η_j into vectors $\mathbf{\Delta Y}$, $\mathbf{q}$, and $\boldsymbol{\eta}$, these equations can be written in vector form as

$$\mathbf{\Delta Y} = A\mathbf{q} + \boldsymbol{\eta}, \tag{2.16}$$

with the components of matrix A given by

$$A_{kj} := \Delta\log\tau\, \Delta t \sum_{i=0}^{k} e^{-(k-i)\Delta t/\tau_j}\, \Delta f_i, \quad k = 0, ..., N-1,\ j = 0, ..., M-1. \tag{2.17}$$

The matrix A is known from the prescribed spectrum of time scales τ_i and the forcings Δf_i. Considering $\boldsymbol{\eta}$ as a fitting error, in principle one can apply standard linear methods to solve eq. (2.16) for the desired spectrum by minimizing

$$\min_{\mathbf{q}_\eta} ||A\mathbf{q}_\eta - \mathbf{\Delta Y}||^2, \tag{2.18}$$

where $||.||$ denotes the Euclidean norm, i.e. $||\mathbf{x}|| = \sqrt{\sum_i x_i^2}$. Here I denoted the spectrum as $\mathbf{q}_{\boldsymbol{\eta}}$ instead of $\mathbf{q}$ to emphasize that the spectrum found in this way can only be an approximation to the original $\mathbf{q}$ depending on the noise present in the data.

Unfortunately, it turns out that solving (2.18) is not a trivial task. The first difficulty is that the finite information provided by the data makes the problem underdetermined:

ideally one wants to obtain a spectrum $q(\tau)$ defined for $\tau \in [0, +\infty[$, but the data $\mathbf{\Delta Y}$ is discrete and covers only a limited time span. However, the most serious issue in identifying $\chi(t)$ arises because (2.9) is a special case of a Fredholm equation of the first kind, where the quest for the integral kernel is well-known to be an ill-posed problem (see e.g. Bertero, 1989, Hansen, 1992). This basically means that any solution $\mathbf{q}_\eta$ of (2.18) obtained via classical numerical methods such as LU or Cholesky decomposition will be extremely sensitive to even small errors in the data (Hansen, 1992). Therefore, to solve (2.18) for the spectrum $\mathbf{q}_\eta$ I invoke regularization.

2.3.3 *Regularized solution*

As discussed above, the problem of solving (2.18) must be expected to be underdetermined and ill-posed. Strictly, the underdetermination cannot be overcome. By contrast, the ill-posed nature of the problem can be treated in the application investigated here. To deal with the ill-posedness, it is useful to perform a Singular Value Decomposition (SVD) of the matrix A:

$$A = U\Sigma V^T \tag{2.19}$$

with $A \in \mathbb{R}^{N\times M}$, $U \in \mathbb{R}^{N\times N}$, $\Sigma \in \mathbb{R}^{N\times M}$, and $V \in \mathbb{R}^{M\times M}$. Σ is a diagonal (not necessarily square) matrix with diagonal entries $\sigma_0 \geq \sigma_1 \geq ... \geq \sigma_{M-1} \geq 0$ known as singular values, and

$$U =: [\mathbf{u}_0, \mathbf{u}_1, ..., \mathbf{u}_{N-1}], \tag{2.20}$$

$$V =: [\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_{M-1}] \tag{2.21}$$

are orthonormal matrices with $\mathbf{u}_0, \mathbf{u}_1, ..., \mathbf{u}_{N-1}$ being the left singular vectors and $\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_{M-1}$ the right singular vectors of A. In practice, assuming that there is more data than prescribed time scales, i.e. $N \geq M$ the singular values σ_i computed numerically are nonzero (see Golub and Van Loan, 1996, section 5.5.8). In this case, (2.18) has the unique solution (see Golub and Van Loan, 1996, Theorem 5.5.1)

$$\mathbf{q}_\eta = \sum_{i=0}^{M-1} \frac{\mathbf{u}_i \bullet \mathbf{\Delta Y}}{\sigma_i} \mathbf{v}_i. \tag{2.22}$$

In practice, when a SVD is applied to a discrete version of a Fredholm equation of the first kind, the components of the singular vectors $\mathbf{v}_i$ and $\mathbf{u}_i$ tend to have more sign changes with increasing index i, as observed by Hansen, 1989. This observation justifies that in the following I dub low-index terms in (2.22) as *low-frequency* contributions, and high-index terms as *high-frequency* contributions.

It is well-known that when applying solution (2.22) one encounters certain numerical problems. Regularization is a means to handle these problems. These problems arise – even in the absence of noise – as follows. From the Riemann-Lebesgue lemma (see e.g. Groetsch,

1984) it is known that the high-frequency components of the data $\Delta Y(t)$ must approach zero. In the discrete case, by Hansen's observation this means that the projections $\mathbf{u}_i \bullet \mathbf{\Delta Y}$ should approach zero for increasing index values i. But due to machine precision or the noise $\boldsymbol{\eta}$ contained in $\mathbf{\Delta Y}$, numerically the absolute values $|\mathbf{u}_i \bullet \mathbf{\Delta Y}|$ do not approach zero but settle at a certain non-zero level for large i or, in the presence of noise, may even increase. Due to the ill-posedness also the singular values σ_i in the denominator of (2.22) tend to zero so that these high-frequency contributions to $\mathbf{q}_\eta$ are strongly amplified. Hence applying eq. (2.22) naively would not give a stable solution for $\mathbf{q}_\eta$ because its value would depend critically on numerical errors and the noise present in the data.

Regularization remedies this problem by suppressing the problematic high-frequency components. This approach assumes that the main information on the solution is contained in the low-frequency components so that the high-frequency contributions to the sum (2.22) can be ignored. This assumption is consistent with the very nature of ill-posed problems because in such problems information on high frequencies is anyway supressed so that only low-frequency components of the solution are recoverable (Groetsch, 1984, section 1.1).

To perform such filtering, I employ the Tikhonov-Phillips Regularization method (Phillips, 1962, Tikhonov, 1963) – although the method is most famously known as Tikhonov regularization, here I choose this different name in recognition of Phillips earlier work (see Groetsch, 2003). Besides being mathematically well-developed (see e.g. Groetsch, 1984, Engl et al., 1996), the Tikhonov-Phillips Regularization method gives an explicit solution in terms of the SVD expansion, which allows for a clear interpretation of the filtering. In addition, it provides a smooth filtering of the solution, in contrast to the also well-known Truncated Singular Value Decomposition method (Hansen, 1987). For additional regularization methods, see e.g. Bertero, 1989, Bertero et al., 1995 and Palm, 2010.

The standard Tikhonov-Phillips Regularization yields the regularized solution in the simple form (Hansen, 2010, Bertero, 1989)

$$\mathbf{q}_\lambda = \sum_{i=0}^{M-1} f_i(\lambda) \frac{\mathbf{u}_i \bullet \mathbf{\Delta Y}}{\sigma_i} \mathbf{v}_i, \tag{2.23}$$

where the $f_i(\lambda)$ are the filter functions

$$f_i(\lambda) = \frac{\sigma_i^2}{\sigma_i^2 + \lambda}. \tag{2.24}$$

By adding the filter function indeed the high-frequency components are suppressed: With λ properly chosen, at large index i, where $\sigma_i^2 \ll \lambda$, $f_i(\lambda)$ approaches σ_i^2/λ so that the terms under sum sign are proportional to σ_i meaning that the terms for large i do not contribute significantly to the sum. In contrast, for small i, $\lambda \ll \sigma_i^2$ so that $f_i(\lambda)$ is about 1 and the terms under the sum are almost unchanged. In this way the filter function indeed selects only the low-frequency components. Therefore, now the problem boils down to determining λ (see next section). Once λ is determined, the solution $\mathbf{q}_\lambda$ is obtained by eq. (2.23) and the desired linear response function $\chi(t)$ finally follows from (2.15).

2.3.4 *Determining the regularization parameter λ from the noise*

By construction it is clear that $\mathbf{q}_\lambda$ as computed from (2.23) strongly depends on the regularization parameter λ. Accordingly, much effort has been put in developing methods to determine suitable values for λ (e.g. Engl et al., 1996, Hansen, 2010). Of special interest are methods that give solutions converging with decreasing noise level to the "true" solution. One such method known to conform to this condition while uniquely determining the regularization parameter has been proposed by Morozov, 1966. His *discrepancy method* is based on the idea that the solution to the problem allows the data to be recovered with an error of the magnitude of the noise (Groetsch, 1984): Let δ denote an upper bound of the *noise level* $||\boldsymbol{\eta}||$, i.e. $\delta \geq ||\boldsymbol{\eta}||$. Then, λ should be chosen such that the discrepancy matches δ, i.e.

$$||A\mathbf{q}_\lambda - \boldsymbol{\Delta}\mathbf{Y}|| = \delta. \tag{2.25}$$

Groetsch, 1983 motivates the choice of this method by demonstrating that determining λ from (2.25) minimizes a natural choice for an upper bound of the error in the solution given by regularization. Unfortunately, the noise level δ is usually not known. But for the application to the carbon cycle intended here, we are in the fortunate situation that typically data from an unforced experiment (control experiment) are available that allows for an independent estimate of the noise level δ.

A naive way to invoke these data to determine λ would be to take δ essentially as the standard deviation of the control simulation – more precisely: $\delta := \sigma\sqrt{N} = ||\boldsymbol{\Delta}\mathbf{Y}_{ctrl}||$. Technically, to find λ, one may start with a large value for λ and decrease it until the left hand side of (2.25) matches δ (as suggested by Hämarik et al., 2011). Having found λ in this way, the desired solution $\mathbf{q}_\lambda$ is then obtained from (2.23). But this approach is not as straightforward as one may think: Because of the forcing, the noise in the perturbed simulation may have different characteristics from that in the control simulation. Therefore in the following I devise a method how to account for this problem.

Formally in (2.16) $\boldsymbol{\Delta}\mathbf{Y}$ is split into a "clean" part and noise $\boldsymbol{\eta}$. Entering this into eq. (2.22) gives

$$\mathbf{q}_\eta = \sum_{i=0}^{M-1} \left(\frac{\mathbf{u}_i \bullet A\mathbf{q}}{\sigma_i}\mathbf{v}_i + \frac{\mathbf{u}_i \bullet \boldsymbol{\eta}}{\sigma_i}\mathbf{v}_i \right). \tag{2.26}$$

Accordingly, the first term in the sum gives the "true" solution $\mathbf{q}$ while the second term gives the noise contribution to $\mathbf{q}_\eta$. As already pointed out when discussing regularization, the "true" solution of ill-posed problems can only be recovered if it is dominated by the projection onto the first singular vectors. This requirement is formally stated by the discrete Picard condition (Hansen, 1990), which demands that the size of the projection coefficients $|\mathbf{u}_i \bullet A\mathbf{q}|$ drops sufficiently fast to zero so that they get smaller than σ_i before σ_i levels off to a finite value because of numerical errors. To find a good estimate for the noise level

δ I use this in the following way. Let i_{max} be the value of the index i where the singular values σ_i start to level off. Assuming that the Picard condition holds, one can infer that

$$\frac{\mathbf{u}_i \bullet \mathbf{\Delta Y}}{\sigma_i} \overset{(2.16)}{=} \frac{\mathbf{u}_i \bullet A\mathbf{q}}{\sigma_i} + \frac{\mathbf{u}_i \bullet \boldsymbol{\eta}}{\sigma_i} \approx \frac{\mathbf{u}_i \bullet \boldsymbol{\eta}}{\sigma_i} \quad \text{for } i > i_{max}. \tag{2.27}$$

This conclusion follows because for $i > i_{max}$ the first term after the equal sign has already dropped towards zero because of the Picard condition, while the projections of the noise $\mathbf{u}_i \bullet \boldsymbol{\eta}$ in the next term are finite, meaning that the first term is vanishing in relation to the second term. Therefore

$$\mathbf{u}_i \bullet \mathbf{\Delta Y} \approx \mathbf{u}_i \bullet \boldsymbol{\eta} \quad \text{for } i > i_{max}. \tag{2.28}$$

This equation determines the high-frequency components of the noise $\boldsymbol{\eta}$. It remains to determine also the low-frequency components to obtain an estimate for δ.

For this purpose, I take advantage of the fact that in CMIP-type experiments one has always a control simulation available. In the control experiment, the simulation is performed for the same conditions as the perturbed simulation, with the only difference that the forcing $\mathbf{\Delta f}$ is zero so that the resulting $\mathbf{\Delta Y}_{ctrl}$ can be understood as pure noise; therefore I write $\boldsymbol{\eta}_{ctrl} := \mathbf{\Delta Y}_{ctrl}$. While in the forced experiment the low-frequency noise is obscured by the low-frequency response induced by the forcing, the low-frequency part of the control simulation data can to first order be expected to give an estimate of the low-frequency noise present in the forced simulation. Nevertheless, it is clear that due to the forcing the spectral characteristics of noise may be different in the forced and unforced experiments. More precisely, the spectrum of noise may differ in its *overall level* and *spectral distribution* (i.e. the "shape" of the spectrum). In the following, I account for a possible difference in the overall level. However, I will assume that the spectral distribution is approximately the same for $\boldsymbol{\eta}_{ctrl}$ and $\boldsymbol{\eta}$; I call this the *spectral similarity assumption.*

After these considerations, λ can be determined as follows: Take i_{max} as the last index i before the plateau $\sigma_i \approx 0$. This i_{max} distinguishes high-frequency ($i > i_{max}$) from low-frequency ($i \leq i_{max}$) components. Then

$$z := || \left[\mathbf{u}_{i_{max}+1} \bullet \mathbf{\Delta Y}, ..., \mathbf{u}_{M-1} \bullet \mathbf{\Delta Y}\right]^T ||, \tag{2.29}$$

$$z_{ctrl} := || \left[\mathbf{u}_{i_{max}+1} \bullet \boldsymbol{\eta}_{ctrl}, ..., \mathbf{u}_{M-1} \bullet \boldsymbol{\eta}_{ctrl}\right]^T || \tag{2.30}$$

are the levels of high-frequency noise in the perturbed (see (2.28)) and in the control simulation, respectively. I now scale the spectral components of $\boldsymbol{\eta}_{ctrl}$ so that its high-frequency level matches the high-frequency level of $\mathbf{\Delta Y}$:

$$\boldsymbol{\eta}' := \frac{z}{z_{ctrl}} \, \boldsymbol{\eta}_{ctrl}. \tag{2.31}$$

In this way, the magnitude of the high-frequency components of $\boldsymbol{\eta}'$ matches that of $\mathbf{\Delta Y}$, and because of eq. (2.28) also that of $\boldsymbol{\eta}$. On the other hand, the spectral distribution of $\boldsymbol{\eta}'$ is the same as for $\boldsymbol{\eta}_{ctrl}$, and by the spectral similarity assumption approximately the same as

for $\boldsymbol{\eta}$. Because $\boldsymbol{\eta}'$ and $\boldsymbol{\eta}$ have similar spectral distributions, the fact that the magnitude of the high-frequency components of $\boldsymbol{\eta}'$ matches that of $\boldsymbol{\eta}$ implies that also the magnitude of their low-frequency components matches. Therefore, $\boldsymbol{\eta}'$ can be seen as an estimate of the noise $\boldsymbol{\eta}$ in the perturbed system not only at high but also at low frequencies. Hence this corrected noise vector $\boldsymbol{\eta}'$ can be used to obtain an estimate of the noise level of the perturbed system by setting

$$\delta := ||\boldsymbol{\eta}'||. \tag{2.32}$$

Compared to taking for δ simply the noise level from the unperturbed simulation (as was insinuated above), taking it in this scaled way assures that the high-frequency components are consistent with the Picard condition that must hold for $\mathbf{q}$ to be recoverable from the ill-posed problem tackled here. Having determined δ, λ can now be computed from (2.25) as described above, from which the $\mathbf{q}$ follows (eq. (2.23)) and hence $\chi(t)$ (eq. (2.15)).

2.3.5 *Additional noise level adjustment in the presence of a monotonicity constraint*

For the application to the land carbon cycle, certain response functions $\chi(t)$ are known to decrease monotonically to zero (see Appendix A.4). In attempts to recover such response functions by employing the noise level adjustment described in the previous section, it may turn out that the numerically found response function fails to be monotonic. There may be several reasons for this failure (strong nonlinearities, signal too obscured by noise, etc.). But one additional reason may be that the low-frequency level of the noise was not properly estimated by assuming that the spectral distribution in the unperturbed simulation reflects the distribution in the perturbed simulation. For such cases one may try to improve the result by further adjustment of the low-frequency noise level to obtain a more reasonable result.

The idea is to adjust the low-frequency components of noise independently of the high-frequency components iteratively until the solution obeys the monotonicity constraint. To understand how to do so, several things must be explained:
1. A sufficient condition for $\chi_\lambda(t)$ being monotonic is that all components q_i have the same sign (see Appendix A.5). Therefore, starting out from a numerical solution for $\chi_\lambda(t)$, it would develop towards monotonicity if one could come up with a sequence of vectors $\mathbf{q}_\lambda$ having less and less sign changes.
2. From eq. (2.23) it is seen that because of Hansen's observation explained in section 2.3.3, that singular vectors $\mathbf{v}_i$ are less noisy for lower i, $\mathbf{q}_\lambda$ has fewer sign changes the fewer $\mathbf{v}_i$ contribute to the sum.
3. As seen from (2.23) and (2.24), this is the case the more components the filter function is suppressing, i.e. the larger the value of λ.
4. To obtain larger values of λ, one sees from the discrepancy method (2.25) that one has to increase δ. The proof for this can be found in Groetsch, 1984 (Theorem 3.3.1) but it can also be made plausible as follows: starting from $\lambda = 0$, $\mathbf{q}_\lambda = \mathbf{q}_\eta$, which is the solution of the minimization problem (2.18). Hence, for $\lambda = 0$ the discrepancy in the left-hand side

of eq. (2.25) is minimal. By increasing λ one decreases all components of $\mathbf{q}_\lambda$ (eq. (2.23)), increasing thereby the discrepancy.
5. Following the reasoning of the previous section, in order to obtain a larger value for δ one must increase the noise level $||\boldsymbol{\eta}'||$ (compare eq. (2.32)). In doing so, one must keep the high-frequency components of $||\boldsymbol{\eta}'||$ unchanged because they must keep matching the level of the high-frequency components of the noise in the perturbed simulation $\boldsymbol{\eta}$ (given by eq. (2.28)). Hence, to increase δ one sees from (2.32) that this is achieved by scaling up the low-frequency components of $||\boldsymbol{\eta}'||$.

Summarizing these considerations, we have to increase the level of low-frequency contributions to $\boldsymbol{\eta}'$ to develop a given solution for $\chi_\lambda(t)$ towards monotonicity.

This leads to the overall algorithm listed in Fig. 2.2. The first five steps have been already explained at the end of section 2.3.4. To account for monotonicity the additional step 6 combined with the loop back to step 4 has to be iteratively executed. To enhance the low-frequency noise level as explained above, I calculate in step 6 a new noise vector $\boldsymbol{\eta}_{new}$ by keeping the high-frequency part from $\boldsymbol{\eta}$ and enhancing its low-frequency components by a factor $c > 1$. Then I recompute $\chi(t)$ from steps 4 and 5 and once more check for monotonicity.

2.4 APPLICABILITY IN THE PRESENCE OF NOISE AND NONLINEARITIES

In application to real data the presence of noise and nonlinearities may complicate the recovery of linear response functions. Therefore in the present section I analyze by using artificial data generated from a toy model the robustness of the RFI method in the presence of such complications and how they can be detected. Insofar this section prepares for the application of the RFI method to the identification of the linear response functions in chapters 3 and 4.

2.4.1 *Toy model and artificial experiments*

As toy model I take

$$\frac{d}{dt}\mathbf{x}(t) = M\mathbf{x}(t) + f(t)\mathbf{a} + \mathbf{n}(t). \tag{2.33}$$

Here the matrix M describes the relaxation of the unperturbed model. The second right-hand side term represents the deterministic forcing constructed from the forcing strength $f(t)$ and the coupling vector $\mathbf{a}$. Additionally, the system is perturbed by the stochastic forcing $\mathbf{n}(t)$, which for simplicity is assumed to be white noise. To make the relation to the carbon cycle considered in the next chapter, the components of $\mathbf{x}$ may be understood as the carbon stored in plant tissues and soils at the different locations worldwide, so that the

1. Take $\boldsymbol{\eta}_{ctrl} := \boldsymbol{\Delta}\mathbf{Y}_{ctrl}$ from the control simulation.
2. Determine i_{max} as the last i before the plateau $\sigma_i \approx 0$.
3. Define

$$z := || \left[\mathbf{u}_{i_{max}+1} \bullet \boldsymbol{\Delta}\mathbf{Y}, ..., \mathbf{u}_{M-1} \bullet \boldsymbol{\Delta}\mathbf{Y}\right]^T ||,$$
$$z_{ctrl} := || \left[\mathbf{u}_{i_{max}+1} \bullet \boldsymbol{\eta}_{ctrl}, ..., \mathbf{u}_{M-1} \bullet \boldsymbol{\eta}_{ctrl}\right]^T ||.$$

   Set

$$\boldsymbol{\eta}' := \frac{z}{z_{ctrl}}\, \boldsymbol{\eta}_{ctrl} \qquad \text{(spectral similarity assumption).}$$

4. Set $\delta := ||\boldsymbol{\eta}'||$, solve (2.25) for λ, and obtain $\mathbf{q}_\lambda$ from (2.23).
5. Compute $\chi(t)$ from (2.15).

This is the final result except a monotonicity should be accounted for. In that case the algorithm proceeds as follows:

6. Check if the resulting $\chi(t)$ decays monotonically to zero. If so, we are done. Else, enhance the low-frequency noise level by setting

$$\boldsymbol{\eta}_{new} := c \sum_{i=0}^{i_{max}} \mathbf{u}_i \bullet \boldsymbol{\eta}' \mathbf{u}_i + \sum_{i=i_{max}+1}^{M-1} \mathbf{u}_i \bullet \boldsymbol{\eta}' \mathbf{u}_i,$$

   where c is some value larger than 1. Then set $\boldsymbol{\eta}' := \boldsymbol{\eta}_{new}$ and repeat calculations starting from step 4.

Figure 2.2 : Final RFI algorithm (see text for details).

observable $Y(t) := \sum_i x_i(t)$ is the analogue of globally stored land carbon. The solution of the system is

$$\mathbf{x}(t) = \int_0^t e^{M(t-s)}\mathbf{a}f(s)ds + \int_0^t e^{M(t-s)}\mathbf{n}(s)ds. \tag{2.34}$$

Since almost every linear system can be diagonalized, I assume from the outset M to be diagonal with eigenvalues $-1/\tau_i^*$, the τ_i^* being the relaxation time scales. Then

$$Y(t) = \int_0^t \chi^*(t-s)f(s)ds + \eta^*(t) \tag{2.35}$$

with the linear response function χ^* and the noise term $\eta^*(t)$ given by

$$\chi^*(t) = \sum_{i=0}^{D-1} a_i e^{-t/\tau_i^*}, \tag{2.36}$$

$$\eta^*(t) = \sum_{i=0}^{D-1} \int_0^t e^{-(t-s)/\tau_i^*} n_i(s)ds, \tag{2.37}$$

where D is the dimension of the system.

To complete the description of the toy model one has to specify its parameters. For the dimension D I take 70 and the time scales are assumed to be distributed logarithmically between $\tau^*_{min} = 0.01$ and $\tau^*_{max} = 1000$, i.e. $\tau^*_i = 0.01 \times 10^{i\Delta log\tau}$ with $\Delta log\tau = (\log_{10} 10^3 - \log_{10} 10^{-2})/70$. The distribution of the components of the coupling vector is adapted from the log-normal rate distribution found by Forney and Rothman, 2012a for the decomposition of soils:

$$a_i = \frac{1}{\tau^*_i \sigma\sqrt{2\pi}} exp\left(-\frac{(\ln \tau^*_i - \mu)^2}{2\sigma^2}\right), \tag{2.38}$$

with μ and σ chosen so that the peak time scale is around $\tau = 5$ and the limits of the log-normal distribution are approximately within $\tau = 0.1$ and $\tau = 200$ (see "true" spectrum in Fig. 2.3). The components of $\mathbf{n}$ are are taken as uncorrelated, i.e. $\langle n_i(0) n_j(t)\rangle = \xi\delta_{ij}\delta(t)$, with standard deviation ξ being chosen differently in different experiments.

In my experiments I explore how $Y(t)$ behaves as a function of the forcing $f(t)$. To this end, I choose a forcing function $f(t)$ (see Table 2.1). The most obvious way to perform the toy model experiments would be to integrate (2.33). But to have a better control over the noise it is for our purpose more appropriate to use the analytical solution (2.35) - (2.37). Hence, I numerically integrate eq. (2.35), using the representation (2.36) and (2.37). The data from these experiments are then used to investigate the performance of the RFI algorithm to recover χ^*. Since all a_i values are non-negative, the response function (2.36) is monotonic, so that I apply the extended version of the algorithm (see Fig. 2.2 including step 6). In all experiments I generate $N = 140$ data points, to have a time series of similar length as in the C^{4}MIP simulations analyzed in the next chapter (140 years, one value for each year). To apply the RFI method also the noise from an associated control simulation is needed. This is obtained from eq. (2.37) by using another realization $n_i(t)$ of white noise for each system dimension i.

2.4.2 *Choice of parameters for the RFI method*

To apply the RFI method, I choose $M = 30$ time scales for the recovery of χ^*. Using $\tau_{min} = 0.1$ and $\tau_{max} = 10^5$ I distribute the spectrum of time scales according to eq. (2.13). These parameters are also used for the application on the carbon cycle in the next chapters and for the comparison with previous methods in section 2.5.

2.4.3 *Ideal conditions*

To gain trust in the numerics of my implementation of the RFI method I present in this section a technical test considering conditions under which it is known that the linear response function should be quite perfectly recoverable. Such ideal conditions are characterized by perfect linearity and absence of noise. Hence I use the presented toy model

(which is anyway linear) in the absence of noise ($\mathbf{n} = \mathbf{0}$) for this test. Actually, this will not be a full test of the algorithm, but only of the implementation of its basic apparatus (sections 2.3.2 and 2.3.3) culminating in eq. (2.23) since in the absence of noise the method to determine the regularization parameter λ (sections 2.3.4 and 2.3.5) is not applicable. One might think that in the absence of noise one could use eq. (2.22) to determine the linear response function, but even under such ideal conditions the ill-posedness of the problem calls for regularization to suppress the *numerical noise* that prevents from obtaining a sensible solution from (2.22) (see discussion in the paragraph after eq. (2.22)). But choosing the extremely small value of $\lambda = 10^{-8}$ for the regularization parameter when evaluating (2.23) is sufficient for this technical test.

Fig. 2.3c shows the response of the noiseless toy model to the forcings shown in Fig. 2.1, i.e. I performed the experiments listed in Table 2.1, although the differences between the "bgc" and "rad" setups do not play a role here and for the present test the control simulation is not needed.

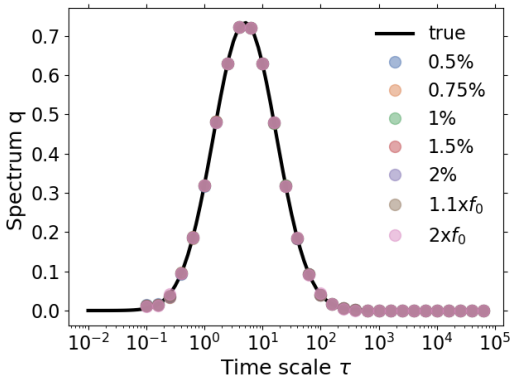
Applying eq. (2.23) to the simulation data gives the spectrum $\mathbf{q}_\lambda$ shown in Fig. 2.3a. Here, I derived the spectrum $\mathbf{q}_\lambda$ for each experiment separately, although in the figure only single dots are seen, because all results coincide so closely and are almost indistinguishable from the "true" solution $\mathbf{q}^*$ as was expected for this ideal case. The next Fig. 2.3b shows the response function obtained from the spectra $\mathbf{q}_\lambda$ using eq. (2.15). Obviously from Fig. 2.3a the "true" response function is reconstructed perfectly from whatever experiment used. As a final test I predict using in eq. (2.1) the response function obtained from the 1% experiment the responses of other experiments. And indeed, these predicted responses are indistinguishable from the responses obtained directly from the experiments (see Fig. 2.3c). This latter result demonstrates perfect robustness of the numerical approach to recover the responses in this ideal case.

2.4.4 *First complication: noise*

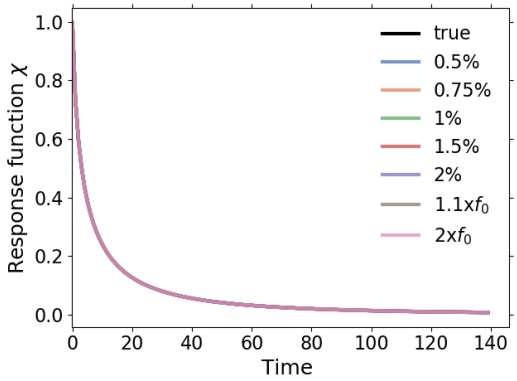
The presence of noise may severely hinder the detailed recovery of $\chi_*(t)$ due to the ill-posed nature of the problem. As already discussed in 2.3.3, small singular values of the matrix A tend to amplify the noise terms associated to the high-frequency singular vectors in the expansion (2.26). Tikhonov-Phillips regularization filters out these terms via (2.23) and (2.24). As a consequence, the accuracy of the solution $\mathbf{q}_\lambda$ tends to decrease as the noise level increases. How severely the accuracy of $\mathbf{q}_\lambda$ decreases depends on the smoothness of the "true" spectrum $\mathbf{q}^*$: The fewer terms one needs to describe $\mathbf{q}^*$ (i.e. the smoother it is), the higher is the level of noise at which it is still possible to recover $\mathbf{q}^*$ with acceptable accuracy. This will get more clear from the Picard plots introduced below when illustrating the effect of noise using experiments with the toy model.

In order to demonstrate the effect of the addition of noise on the quality of the derived $\chi(t)$, I define a relative error for the prediction of the responses from different experiments.

(a) Spectrum of time-scales q



(b) Response function χ



(c) Prediction from 1%

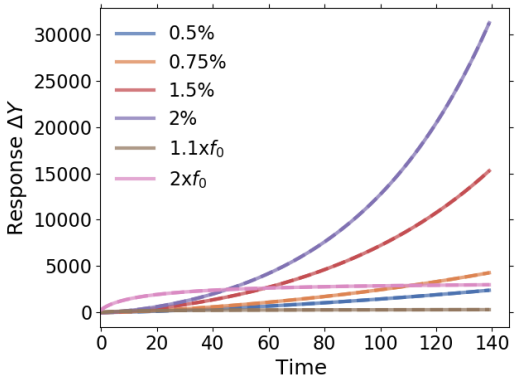


Figure 2.3 : Demonstration of robust recovery for noise-free data from the toy model: (a) recovered $\mathbf{q}_\lambda$; (b) recovered $\chi(t)$; and (c) original responses (dashed lines) and predictions (continuous lines) using $\chi_\lambda(t)$ derived from the 1% experiment. Reconstructed values are almost indistinguishable from original data. For plotting the "true" spectrum of the toy model in subfigure (a) I used the relation $\mathbf{q}^* = \mathbf{a}/\Delta \log_{10} \tau$, which can be obtained by comparing eq. (2.36) with eq. (2.15). Since from the discrete spectrum the response function and the response may be obtained for any time t, the spectrum is plotted as dots while the response function and response are plotted as continuous lines.

Consider a particular experiment – which is in our case the 1% experiment – from which we have obtained by the RFI method the response function, which I call here $\chi^0(t)$. The relative error for the prediction of the response from an experiment "k" by the recovered $\chi^0(t)$ via the convolution (2.1) is

$$\varepsilon_k^0 := \frac{||\mathbf{\Delta Y}^k - \boldsymbol{\chi}^0 \star \mathbf{\Delta f}^k||}{||\mathbf{\Delta Y}^k||}, \tag{2.39}$$

where $\star$ stands for the discrete form of the convolution operation (2.1) used to predict the responses. In the following I denote ε_k^0 the *prediction error* for the experiment "k". To measure the quality of the prediction across multiple experiments I define also the *mean prediction error*

$$\varepsilon^0 := \frac{1}{K}\sum_{k=1}^{K}\varepsilon_k^0, \tag{2.40}$$

where K is the number of predicted responses. The reader may wonder why I quantify the quality of the recovery only indirectly from the responses found in different experiments and not directly from the recovery of $\chi(t)$. The reason is that in real applications the "true" $\chi(t)$ is not known but the responses are. The reliability of this indirect measure for the quality of the recovery is discussed in section 2.5.

To study how the quality of the recovery depends on the noise level I introduce the signal-to-noise ratio (SNR) of the response data from a perturbation experiment as

$$SNR := \frac{||\mathbf{\Delta Y}||}{\delta}, \tag{2.41}$$

where δ is the final noise level estimate obtained by the RFI method, as described in section 2.3.3 (see eq. (2.32)).

To demonstrate the dependence of the mean prediction error (2.40) on the SNR, I performed 1% experiments using different noise levels. The resulting dependence is shown in Fig. 2.4. As expected, for a small error a sufficiently large SNR is needed, i.e. a good recovery may be hindered by a too large noise level.

In Fig. 2.5 I demonstrate how the overall noise level adjustment in step 3 of the RFI algorithm (see Fig. 2.2) affects regularization to recover the correct response function. To guarantee that the overall level of the noise spectrum is indeed substantially different in the control and perturbed simulations (so that the adjustment is really needed), I take for the noise in the control simulation a standard deviation ten times smaller than that for the noise in the perturbed simulation. To demonstrate how the adjustment works it is helpful to consider the so-called "Picard plot". This type of plot was originally introduced to analyze the spectral characteristics of an ill-posed problem (see e.g. Hansen, 1992). In Fig. 2.5a I show the Picard plot for data obtained from a 1% simulation with the toy model using a SNR $\approx$ 520 to assure a good recovery. The singular values σ_i decrease to extremely small values as the index i increases. This demonstrates that indeed the problem

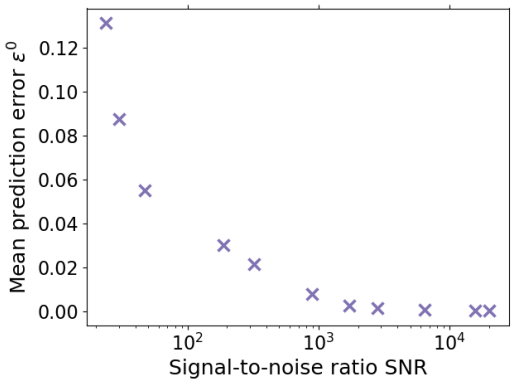


Figure 2.4 : Mean prediction error (2.40) of the recovery when deriving $\chi(t)$ for different values of the SNR. As the SNR increases, the recovery of $\chi(t)$ improves. To illustrate the most general case where $\chi(t)$ is not known to be monotonic I do not apply the monotonicity check (step 6 of Fig. 2.2).

to solve for the response function is ill-posed and therefore regularization is needed for its solution (compare eq. (2.22) with (2.23)). The data labelled by $|\mathbf{u}_i \bullet \boldsymbol{\eta}|$ are the "true" noise coefficients, obtained by subtracting the "clean" response $A\mathbf{q}$, known analytically from the toy model description, from the noisy toy model response $\boldsymbol{\Delta}\mathbf{Y}$. Comparing them to the projection coefficients of the response $|\mathbf{u}_i \bullet \boldsymbol{\Delta}\mathbf{Y}|$ one sees that with exception of the first few coefficients the response is dominated by its noise content. Accordingly, only the information contained in these first few coefficients is recoverable from this ill-posed problem whatever method is used. The data labelled by $|\mathbf{u}_i \bullet \boldsymbol{\eta}_{est}|$ have been added to the Picard plot to demonstrate how the RFI algorithm operates: These data are the projection coefficients of the estimated noise content in the data, where $\boldsymbol{\eta}_{est}$ is the final value of $\boldsymbol{\eta}'$ obtained by the RFI method. Obviously, the RFI algorithm correctly estimates the "true" noise level not only at high frequencies – where it is correct by the noise level adjustment in step 3 of the RFI algorithm (see Fig. 2.2) – but also at low frequencies, where it is predicted from the adjusted low-frequency components of the control simulation (also step 3). Acordingly, in this case the spectral similarity assumption holds and there is no need to further adjust of the noise level (step 6).

How the estimation of the noise in the data and the resulting regularization affects the projection coefficients of the spectrum $\mathbf{q}$ can be seen in Fig. 2.5b: Only those few coefficients not dominated by noise contribute to the regularized solution. In this case these few coefficients selected by determinining the regularization parameter λ from the noise level are sufficient for an almost perfect recovery of the response function, as seen in Fig. 2.5c.

It is important to note that in the situation of Fig. 2.5 where the overall noise level differs considerably in the control and in the perturbed simulation, a naive noise estimate taken from the control simulation without the adjustment in step 3 (as first suggested in section 2.3.4) would severely underestimate the noise actually in the data. This would in turn lead to

an underestimation of the regularization parameter (see Groetsch, 1984, Theorem 3.3.1). As a result, the wrong filtering by regularization would leave projection coefficients dominated by noise in the solution, likely leading to large errors in the recovered response function. This example therefore demonstrates the relevance of the noise adjustment in step 3.

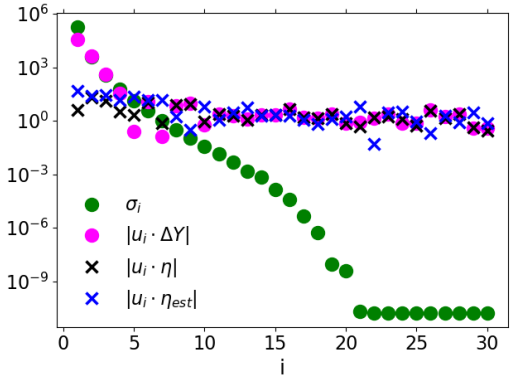
Finally in this section, I demonstrate that by accounting for monotonicity of the linear response function one may obtain a better estimate of the low-frequency components of the noise whereby the recovery of the response function is improved. In Fig. 2.6 I plot results from toy model simulations where the spectral similarity assumption does not hold. This was achieved by artificially enhancing the low-frequency components of the noise $\eta^*(t)$ in eq. (2.34). The top row plots show the results from the recovery when the additional noise level adjustment was not used. Because the spectral similarity assumption does not hold, the estimated low-frequency components of the noise $|\mathbf{u}_i \bullet \boldsymbol{\eta}_{est}|$ do not match those of the "true" noise $|\mathbf{u}_i \bullet \boldsymbol{\eta}|$ (subfigure (a_1)). Ideally, only those four projection coefficients of the data $|\mathbf{u}_i \bullet \boldsymbol{\Delta}\mathbf{Y}|$ which are larger than the projection coefficients of the "true" noise $|\mathbf{u}_i \bullet \boldsymbol{\eta}|$ should contribute to the recovered response function. Instead, as seen in subfigure (b_1), the coefficients with index between $i = 5$ and 8 give the dominant contributions because they are larger than the estimated noise coefficients $|\mathbf{u}_i \bullet \boldsymbol{\eta}_{est}|$ (compare subfigure (a_1)). Therefore, the recovery of the response function is poor (subfigure (c_1)). But since in this case the low-frequency components of noise are such that the recovered response function is non-monotonic although the "true" response function is known to be monotonic, one may further adjust the noise level to improve the results.

This further adjustment is the purpose of step 6 of the RFI algorithm (see Fig. 2.2). Its effect is demonstrated by the second-row plots of Fig. 2.6: The estimated noise components match now better the "true" noise components that had been underestimated in the first row (compare subfigures (a_2) and (a_1)) so that only those four components that carry information (compare in subfigure (a_2) the projections $|\mathbf{u}_i \bullet \boldsymbol{\Delta}\mathbf{Y}|$ for low index i with $|\mathbf{u}_i \bullet \boldsymbol{\eta}|$) survive the regularization (subfigure (b_2)). As a result, the recovery error of the response function has considerably decreased (subfigure (c_2)).

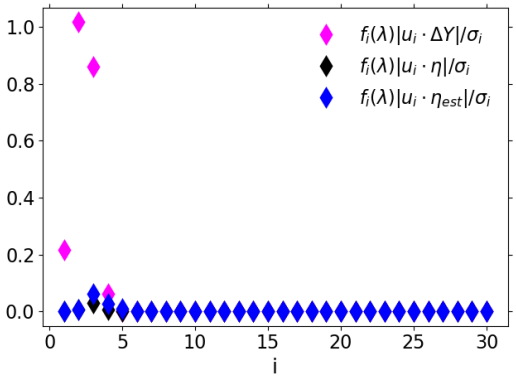
2.4.5 *Second complication: nonlinearity*

The second difficulty in recovering the linear response function $\chi(t)$ from a perturbation experiment may arise from nonlinearities present in the considered system. Generally it must be suspected that nonlinearities are present so that they should not hurt as long as they are small. And indeed, from the viewpoint of regularization, contributions from nonlinearities can be considered as an additional noise so that in principle they can also be filtered out. But as with noise, when getting stronger they cause a deterioration of the recovery of the response function. In the following, I show this more formally and discuss in detail how the RFI algorithm behaves in the presence of nonlinearities.

(a) Picard plot



(b) Coefficients of solution (2.23)



(c) Linear response function

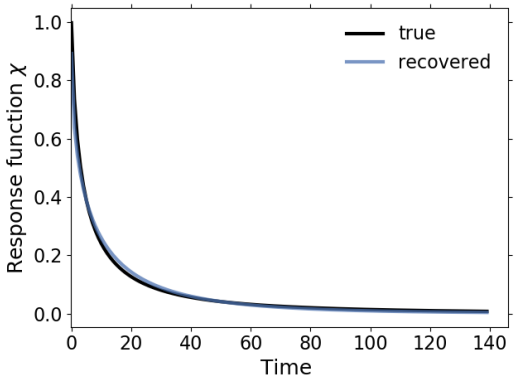


Figure 2.5 : Demonstration of the operation of the RFI algorithm in the presence of noise using toy model data from a 1% and a control experiment. To demonstrate the relevance of the noise level adjustment (step 3 from Fig. 2.2), the standard deviation of the noise in the control simulation was taken ten times smaller than that for the noise in the perturbed simulation. (a) Picard plot showing the singular values σ_i and the projection coefficients of the data $|\mathbf{u}_i \bullet \mathbf{\Delta Y}|$, the "true" noise $|\mathbf{u}_i \bullet \boldsymbol{\eta}|$, and the final noise estimate $|\mathbf{u}_i \bullet \boldsymbol{\eta}_{est}|$; (b) coefficients of regularized solution (2.23); (c) "true" and recovered linear response functions. Since the RFI algorithm correctly adjusted the noise level to the "true" noise in the data, the resulting regularized solution has contributions only from the first few projection coefficients which are not completely obscured by noise. Overall, the recovery is almost perfect, because the SNR (chosen as about 650) is still sufficiently good and because the noise was chosen to conform with the spectral similarity assumption. Because the noise level adjustment (step 3 from Fig. 2.2) already gave a good estimate to the "true" noise in the data, no monotonicity check was needed (step 6 from Fig. 2.2).

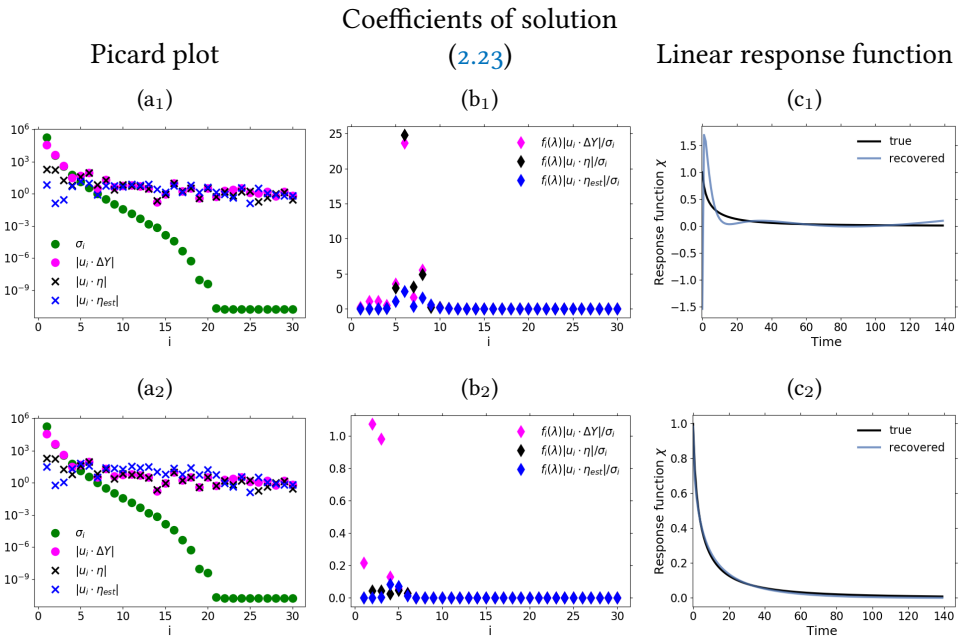


Figure 2.6 : Demonstration of the additional noise level adjustment in the presence of a monotonicity constraint using toy model data from a 1% and a control experiment: (a) Picard plot; (b) coefficients of regularized solution (2.23) and (c) recovered linear response function. All figures are based on the same toy model simulations using a SNR = 1189. To demonstrate the effect of the noise level adjustment the spectral similarity assumption is broken by artificially increasing the low-frequency components of the noise in the simulations. The plots in the first row show the results from the RFI algorithm in the absence of additional noise level adjustment (step 6 in Fig. 2.2). Although the "true" response function of the toy model is monotonic, the response function recovered by the RFI algorithm is non-monotonic (last figure in the first row). But if the noise adjustment is switched on (second row), the response function is correctly recovered as monotonic (last figure in the second row). For more details see text.

To understand how contributions from nonlinearities affect the recovery of the response function I write the nonlinear terms in eq. (2.8) collectively as $\widetilde{\eta}(t)$. This formally gives

$$\mathbf{\Delta Y} = A\mathbf{q} + \boldsymbol{\eta} + \widetilde{\boldsymbol{\eta}} \tag{2.42}$$

instead of eq. (2.16). Plugging this into (2.23) the spectrum is obtained as

$$\mathbf{q}_\lambda = \sum_{i=0}^{M-1} f_i(\lambda) \left(\frac{\mathbf{u}_i \bullet A\mathbf{q}}{\sigma_i}\mathbf{v}_i + \frac{\mathbf{u}_i \bullet (\boldsymbol{\eta} + \widetilde{\boldsymbol{\eta}})}{\sigma_i}\mathbf{v}_i \right). \tag{2.43}$$

Accordingly, the nonlinear contributions can be understood as an additional noise in the spectrum $\mathbf{q}_\lambda$ so that the theory of regularization fully applies when replacing $\boldsymbol{\eta}$ by the *combined noise* $\boldsymbol{\eta} + \widetilde{\boldsymbol{\eta}}$. Hence, as in their absence, nonlinearities do not prevent the application of regularization as long as the signal is not buried under this combined noise.

But for the RFI algorithm to give good results a second condition is that the contributions from $\widetilde{\boldsymbol{\eta}}$ must not be large compared to those from $\boldsymbol{\eta}$. To understand this, one must realize that the response and with it the nonlinear contributions $\widetilde{\boldsymbol{\eta}}$ are dominated by low-frequency components because of the low-frequency nature of the forcing for the problems of interest (for instance in %-experiments). The RFI algorithm uses an estimate for the noise level in the perturbation experiment obtained from the control simulation assuming that the spectral distribution is approximately the same in the noise from the control simulation and the noise in the data from the perturbation experiment (spectral similarity assumption; step 3 of Fig. 2.2). But the control simulation does not contain any contributions from nonlinearities because the forcing is zero. Therefore, if in the data from the perturbation experiment the contributions from nonlinearities $\widetilde{\boldsymbol{\eta}}$ are not small compared to those from $\boldsymbol{\eta}$, the spectral similarity assumption does not hold. Since this assumption is at the heart of the RFI algorithm, its breakdown leads to a poor recovery of the linear response function.

All this is demonstrated in the following by toy model simulations. For this purpose, I artificially consider the response of the toy model not in Y but in its nonlinear transform

$$Y_{nonlin}(t) := Y(t) - aY^2(t), \tag{2.44}$$

where the parameter a determines the strength of the nonlinearity. Indeed, in such a way the nonlinearity does not result from nonlinearity of the underlying dynamics (the toy model is linear), but from the way the response is looked at. But this distinction is artificial since in practice a response experiment is an indivisible unity of system and observation so that the origin of the nonlinearity is irrelevant. The particular functional form chosen for $Y_{nonlin}(t)$ mimics the nonlinear effect of saturation encountered for instance in the land carbon sink when atmospheric CO_2 rises to high values. In the following, to demonstrate the effect of nonlinearities, I set the noise level in the toy model experiments to a rather small value in order to have a good SNR in the experiments considered.

In Fig. 2.7 I show by plotting the mean prediction error (see eq. (2.40)) how the recovery of the response function deteriorates as the nonlinearity parameter a increases. To demonstrate

that this is indeed caused by a breakdown of the spectral similarity assumption I plot in addition the ratio $||\widetilde{\boldsymbol{\eta}}||/||\boldsymbol{\eta}||$. It is seen that indeed, as claimed above, the recovery works well only when this ratio is not large, i.e. when the contributions from nonlinearities $\widetilde{\boldsymbol{\eta}}$ are not large compared to those from the noise $\boldsymbol{\eta}$.

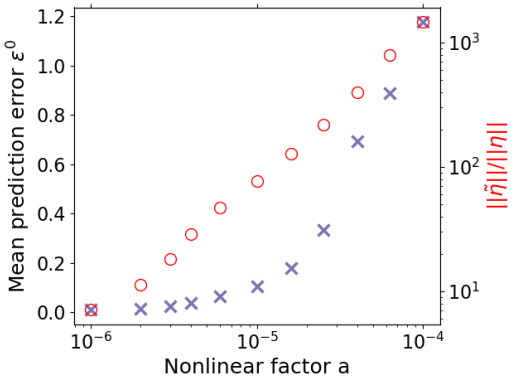


Figure 2.7 : Mean prediction error (2.40) of the recovery when deriving $\chi_\lambda(t)$ for different values of the nonlinearity factor a of the toy model. As a increases, the recovery of $\chi_\lambda(t)$ deteriorates because the level of the contributions from nonlinearities $||\widetilde{\boldsymbol{\eta}}||$ gets large compared to the noise level $||\boldsymbol{\eta}||$; how these terms are computed for the toy model is explained in Appendix A.6. To demonstrate here the pure effect from the breakdown of the spectral similarity assumption, the RFI algorithm is used here without the additional noise level adjustment enforcing monotonicity.

More insight into how nonlinearities affect the recovery is obtained from the more detailed SVD analysis shown in Fig. 2.8. The first row of subfigures was obtained from the toy model assuming a rather small nonlinearity ($a = 10^{-10}$). In the Picard plot (subfigure (a_1)) it is seen that in this case both conditions necessary for a good recovery are met: First, the signal $|\mathbf{u}_i \bullet \boldsymbol{\Delta}\mathbf{Y}|$ is well visible above the combined noise $|\mathbf{u}_i \bullet (\boldsymbol{\eta} + \widetilde{\boldsymbol{\eta}})|$ (see the first four components). Second, in this case $|\mathbf{u}_i \bullet \widetilde{\boldsymbol{\eta}}|/|\mathbf{u}_i \bullet \boldsymbol{\eta}|$ is small over the whole spectrum, i.e. the contributions from $\widetilde{\boldsymbol{\eta}}$ are small compared to those from $\boldsymbol{\eta}$. As explained above, because this second condition is also met, the noise estimate from the RFI algorithm $\boldsymbol{\eta}_{est}$ is a good approximation to the combined noise across all frequencies (compare in the Picard plot $|\mathbf{u}_i \bullet (\boldsymbol{\eta} + \widetilde{\boldsymbol{\eta}})|$ to $|\mathbf{u}_i \bullet \boldsymbol{\eta}_{est}|$). As a result, the four components selected by the regularization for the recovered solution (subfigure (b_1)) are precisely those dominated by the signal (compare $f_i(\lambda)|\mathbf{u}_i \bullet \boldsymbol{\Delta}\mathbf{Y}|/\sigma_i$ with $f_i(\lambda)|\mathbf{u}_i \bullet (\boldsymbol{\eta} + \widetilde{\boldsymbol{\eta}})|/\sigma_i$). This example demonstrates that as long as these two conditions are met, small contributions from nonlinearities do not prevent a good recovery of the response function (see subfigure (c_1)).

In the second row of Fig. 2.8, I demonstrate how the violation of the second condition obstructs the recovery. In this case the nonlinearity parameter has been given a larger value ($a = 2.5 \times 10^{-5}$). As a consequence, one sees in the Picard plot that the low-frequency components of the combined noise are enhanced. The first condition is still met: the signal $|\mathbf{u}_i \bullet \boldsymbol{\Delta}\mathbf{Y}|$ is visible above the combined noise $|\mathbf{u}_i \bullet (\boldsymbol{\eta} + \widetilde{\boldsymbol{\eta}})|$ (see the first two components).

But now $|\mathbf{u}_i \bullet \widetilde{\boldsymbol{\eta}}|/|\mathbf{u}_i \bullet \boldsymbol{\eta}|$ gets large at low frequencies, violating the second condition. As explained, the violation of the second condition leads to the breakdown of the spectral similarity assumption. As a result, the RFI algorithm underestimates the combined noise at low frequencies (compare in the Picard plot $|\mathbf{u}_i \bullet (\boldsymbol{\eta} + \widetilde{\boldsymbol{\eta}})|$ to $|\mathbf{u}_i \bullet \boldsymbol{\eta}_{est}|$). Using this wrong noise estimate, regularization selects components for the recovered solution that are to a large extent dominated by the combined noise (see components 3 to 8 in subfigure (b_2)). The result is that the strong low-frequency contributions from nonlinearities deteriorate the recovery of the response function at long time scales (subfigure (c_2)).

In the third row, I demonstrate for this type of nonlinearity that by accounting for monotonicity one can remove from the recovered solution all components dominated by noise. For this purpose, I set the nonlinearity parameter to the same value as for the second row ($a = 2.5 \times 10^{-5}$) but employ the additional noise level adjustment (step 6 of Fig. 2.2), i.e. the low-frequency range of the noise estimate is now automatically adjusted in order to recover a response function that decays monotonically to zero. As seen in the Picard plot, the additional noise level adjustment results in an artificial enhancement of the low-frequency components of the noise estimate, with a large jump separating the low- from the high-frequency range. In this case, such enhancement is able to correctly estimate the largest components of the combined noise (first few components in the Picard plot). As a consequence, regularization correctly selects for the recovered solution only the two first components which are not dominated by noise (subfigure (b_3)). Unfortunately, as seen in subfigure (c_3), these two first components do not contain enough information for a perfect recovery, since the quality improves at long time scales, but deteriorates at short time scales (compare subfigures (c_3) and (c_2)). This is a consequence of how regularization works: It filters out components dominated by noise (or in this case, nonlinearity) at the expense of removing also useful information contained in those components.

Also interesting to note from this SVD analysis is that although in general the presence of nonlinearities cannot be detected from only the two experiments needed for our RFI method, it can be detected in cases where the response function is known to be monotonic but the nonlinearity is such that the recovered response function is non-monotonic. This is shown in the last example above, where strong nonlinearities result in a large jump between the low- and high-frequency components of the noise estimate. Such jump arises because strong nonlinearities cause the response function to be non-monotonic, and this enforces the additional adjustment of the noise estimate by the RFI algorithm. This effect is obviously a result of the particular type of nonlinearity considered for that example. Nevertheless, such jump may be a relevant indication of strong nonlinearities in applications to the land carbon cycle because this type of nonlinearitiy mimics precisely the saturation behaviour observed in the land carbon sink under high values of atmospheric CO_2.

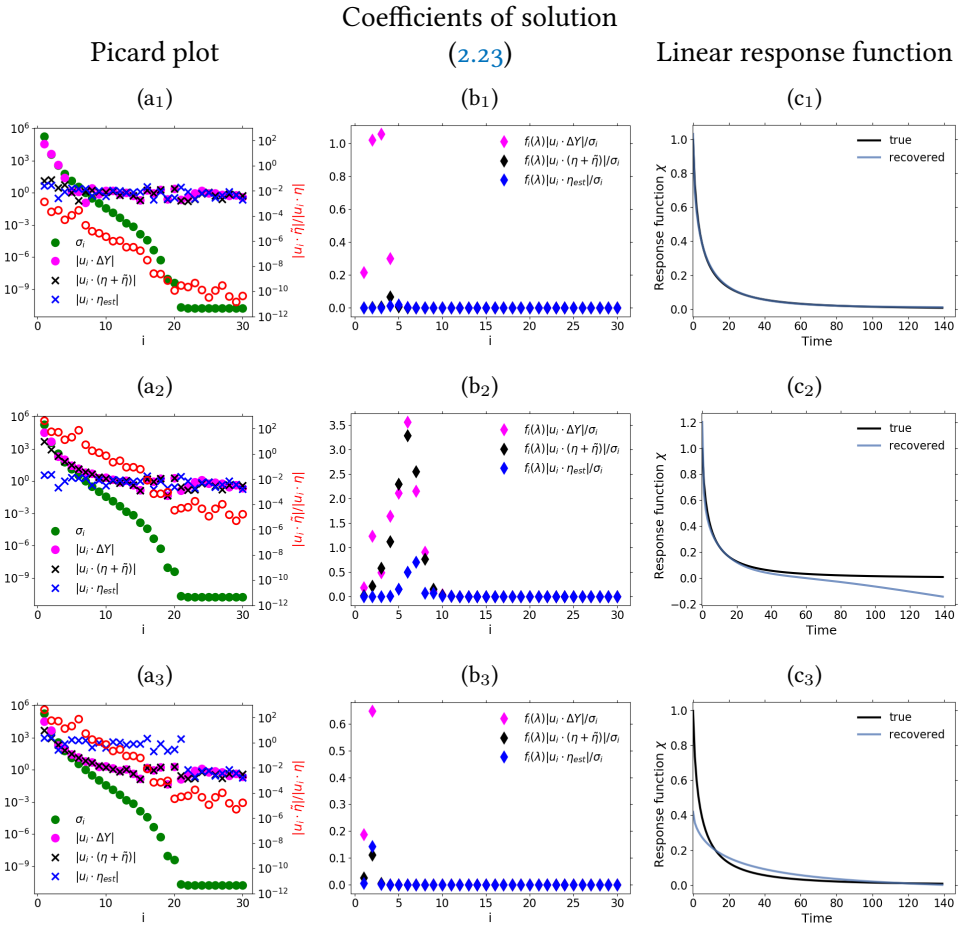


Figure 2.8 : Demonstration of how nonlinearities affect the recovery of the response function: (a) Picard plot; (b) coefficients of regularized solution (2.23) and (c) recovered linear response function. First row: Nonlinearity factor $a = 10^{-10}$ (no monotonicity check); Second row: Nonlinearity factor $a = 2.5 \times 10^{-4}$ (no monotonicity check); Third row: Nonlinearity factor $a = 2.5 \times 10^{-4}$ (with monotonicity check). The noise is overestimated in the low-frequency spectrum in the third row because nonlinearities yield a derived $\chi_\lambda(t)$ that does not obey the monotonicity constraint. As a consequence, the method increases the level of low-frequency components until the monotonicity constraint is obeyed. The failure to obey the monotonicity constraint and consequent large overestimation of noise in this case can be taken as an indication of the presence of nonlinearities in the response. Note that the "true" linear response function in this nonlinear case $a \neq 0$ is obtained analytically from the linear case $a = 0$ via eq. (2.44) (see Appendix A.6). For more details see text.

2.5 COMPARISON WITH PREVIOUS METHODS

As a last test of the quality of the results given by the RFI method before I apply it to the carbon cycle, in this section the RFI method is compared against two existent methods in the literature to identify response functions in the time domain. The comparison is performed for the particular case where the response function is known to be monotonic and also for the more general case where it is not. As a side issue this section reveals also some insight into the relation between the quality of the recovery of $\chi(t)$ as measured by the prediction of responses, and the quality of the recovery of $\chi(t)$ itself.

In climate science, the most commonly used method is to obtain $\chi(t)$ from an impulse response, i.e. the response to a perturbation of Dirac delta-type (e.g. Siegenthaler and Oeschger, 1978, Maier-Reimer and Hasselmann, 1987, Joos and Bruno, 1996, Joos et al., 1996, Thompson and Randerson, 1999, Joos et al., 2013). I call it here *pulse method*. Although this method is conceptually straightforward, in some cases it might not yield satisfactory results. Since the perturbation is only one "pulse", depending on the observable of interest it may give a response with small SNR. As a consequence, the recovered response function may be severely affected by noise. On the other hand, if the strength of the pulse is made large to obtain a good SNR, the linear regime may be exceeded. In this case, the impulse response does not correspond anymore to the linear response function.

The second method consists of deriving the linear response function from a step response, i.e. the response to a Heaviside-type perturbation (e.g. Hasselmann et al., 1993, Ragone et al., 2016, MacMartin and Kravitz, 2016, Lucarini et al., 2017, Van Zalinge et al., 2017 and Aengenheyster et al., 2018). I call it here *step method*. Due to the special form of this "step" perturbation, the linear response function can in principle be derived from

$$\chi(t) = \frac{1}{\Delta f_{step}} \frac{d}{dt} \Delta Y_{step}(t), \tag{2.45}$$

where Δf_{step} is the step perturbation and ΔY_{step} is the corresponding response. Unfortunately, such derivation involves numerical differentiation, which is known to be an ill-posed problem (Anderssen and Bloomfield, 1974, Engl et al., 1996). Because the problem is ill-posed, noise is amplified, potentially resulting in large errors in the derived linear response function.

These two methods therefore share two limitations: First, they require a special perturbation experiment; second, because of noise in the data they might yield a response function with large errors. In principle, the second limitation may be overcome by using instead of a single response the ensemble average over multiple responses. But this comes at the expense of the numerical burden of performing multiple experiments, which is especially large when dealing with complex models such as state-of-the-art Earth System Models.

The main advantages of the RFI method lie precisely in overcoming these two limitations: it recovers the response function from any type of perturbation experiment and automatically filters out the noise by regularization.

For the results of this section, I performed ensembles of 200 simulations with the toy model (see section 2.4.1). Each ensemble member is defined by a realization of the noise $\eta^*(t)$ with a fixed standard deviation (see eq. (2.37)). Each realization was added via eq. (2.35) to three experiments: 1%, step ($2\times f_0$), and pulse ($4\times f_0$). Note that because of the issue with the SNR mentioned above, I had to employ for the pulse experiment twice the forcing strength employed for the step experiment. Further, for each ensemble member an additional realization of the noise was generated to serve as a control simulation to compute the noise estimate for the RFI method (step 1 of Fig. 2.2).

I computed the response function by the pulse and step method as follows. For a pulse experiment the forcing is $\Delta f(t) = a\delta(t)$ with forcing strength a, so that the response is given by

$$\Delta Y_{pulse}(t) = \int_0^t \chi(t-s)a\delta(s)ds = a\chi(t). \tag{2.46}$$

Therefore, for the pulse method I took the response from the pulse experiment and obtained the response function by

$$\chi(t) = \frac{1}{a}\Delta Y_{pulse}(t). \tag{2.47}$$

The recovery by the step method was calculated by taking the response from the step experiment and applying eq. (2.45). The derivative was computed by forward difference.

To obtain comparable results with these two methods, I recovered the response function by the RFI method from the same pulse and step experiments. To compare the quality of the results using also an experiment not decidedly tailored for the identification, I include additionally the recovery from the 1% experiment.

To obtain a quantitative comparison for the quality of the recovery for each method, I define the recovery error

$$\varepsilon^r := \frac{||\boldsymbol{\chi} - \boldsymbol{\chi}^*||}{||\boldsymbol{\chi}^*||}, \tag{2.48}$$

where $\boldsymbol{\chi}$ is the recovered response function and $\boldsymbol{\chi}^*$ is the "true" response function, which is known because we use the toy model. In contrast to the prediction error, that measures the quality of the recovery of $\chi(t)$ by means of the response (see eq. (2.48)), the recovery error ε^r measures the quality of the recovery of $\chi(t)$ itself. Another reason for introducing the recovery error is to compare its results with results from the prediction error. By doing that, we can gain insight into how much the prediction error can be trusted as an indirect measure of the quality of recovery in real applications, where the "true" response function is not known.

First, I compare the pulse and step methods against the full RFI algorithm, i.e. the RFI algorithm taking monotonicity into account (step 6 in Fig. 2.2). Results are shown in Fig. 2.9. In the first row of subfigures, I took for the recovery the ensemble average over the 200 responses for each experiment. For the RFI method, I took the ensemble average over the control simulations as well to estimate the noise (step 1 of Fig. 2.2). As shown in subfigure

(a_1), with this approach all methods recover the response function almost perfectly. The quality of the recovery is quantified by the recovery error in subfigure (b_1). The RFI method shows the smallest values for the step and pulse experiments when compared to the step and pulse methods. Overall, the step method clearly shows the largest value. To quantify the quality of the prediction, I plot in subfigure (c_1) the prediction error (2.39). As seen, values are even smaller than for the recovery error. Overall, we see a similar pattern: the step method again stands out, with other methods showing much smaller error values.

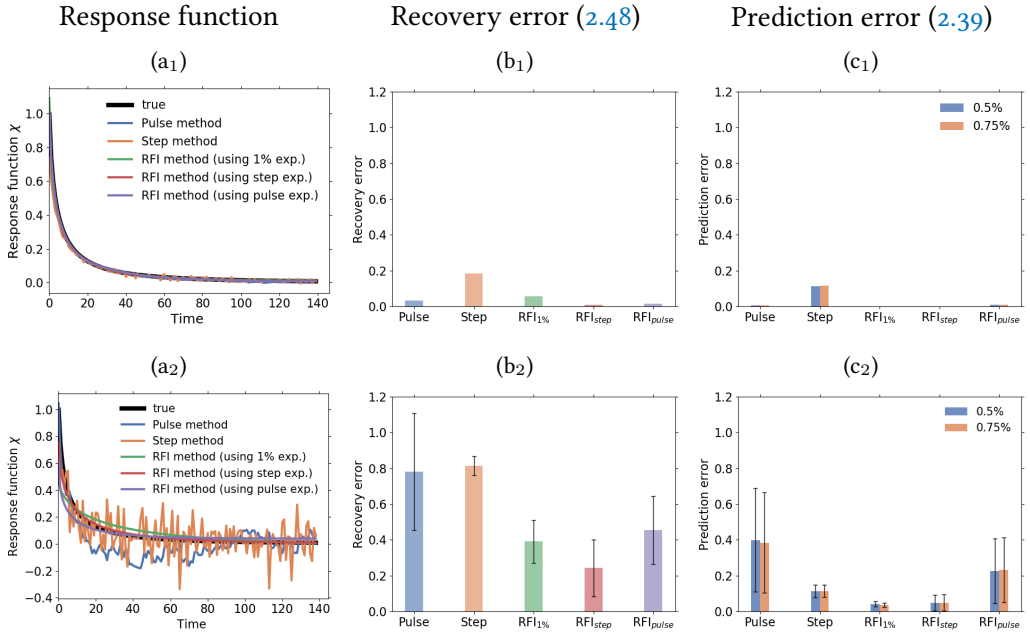


Figure 2.9 : Quality of response function recovery by the full RFI method (including step 6 in Fig. 2.2) in comparison to the pulse and step method. Subscripts at "RFI" indicate the experiment from which the response function was recovered with the RFI method. First row: taking the average over the whole ensemble of toy model experiments for recovery; second row: Performing the recovery for each ensemble member separately. (a_1) Recovered response function; (b_1) recovery error; (c_1) prediction error (2.39); (a_2) example of recovered response function from one ensemble member; (b_2) statistics of recovery error; (c_2) statistics of prediction error (2.39). The prediction error is separately computed for the 0.5% and 0.75% experiments. Taking the ensemble average, all methods perform well (see first row). But taking only one ensemble member, the RFI algorithm gives better recovery and prediction errors than the pulse and step methods when comparing the same responses (see second row).

In the second row, I compare results by taking only a single response for the recovery. Since the quality of the recovery by the different methods may vary depending on the particular noise realization, I again performed 200 simulations to obtain better statistics, but this time deriving the linear response function for each ensemble member separately. Subfigure (a_2) shows an example of recovery for one of the ensemble members. As expected, the recoveries by the pulse and step methods largely deviate from the true response function.

For the pulse method, the large errors result from the low SNR of the pulse response: Even taking twice the forcing strength of the step experiment, the SNR of the pulse response is of order 10^0 against order 10^1 for the step and 1% responses. For the step method, on the other hand, the large errors are not a result of low SNR, but of the noise amplification associated with the ill-posedness of numerical differentiation. In contrast to the recovery by these two methods, because of regularization the recoveries by the RFI method are smoother and visually seem to better fit the true response function. To quantitatively check these results, I plot in subfigure (b_2) for each method the average and standard deviation over the 200 values of the recovery error (one for each ensemble member). The figure shows that indeed the pulse and step methods display the largest average recovery error, with the pulse method having a much larger spread. Such spread is probably related to the low SNR in the response from the pulse experiment. The results from the 1% and pulse experiments by the RFI method are better, showing comparable error magnitudes. The smallest average recovery error is obtained from the RFI method using the step experiment, which is in agreement with conclusions from section 3.3. In subfigure (c_2) I show the average and standard deviation over the 200 values of the prediction error (2.39). The smallest average prediction errors are obtained from the RFI method using the 1% and step experiments. The largest errors are obtained for the pulse method and the RFI method using the pulse experiment. In contrast to the situation for the recovery error, for the prediction error no substantial difference between the two is found. Note also that when comparing recoveries from the same response (i.e. comparing "Pulse" with "RFI$_{pulse}$" and "Step" with "RFI$_{step}$"), the RFI method gives better results than both the pulse and step methods. Another interesting point is that prediction errors for the step method remain approximately unchanged by taking the ensemble mean and a single response (compare "Step" in subfigures (c_1) and (c_2)). Overall, as in the first row, the prediction error shows for each individual method values smaller than the recovery error. But now there is a difference between the plots for the recovery and prediction error: Although the pulse and step methods show the largest averages with values of comparable size for the recovery error, for the prediction error the pulse method has the largest average with a value much larger than the step method.

This difference can be better understood as follows (see MacMartin and Kravitz, 2016 for more details including the influence of the forcing scenario). Because eq. (2.1) is ill-posed, the convolution operator acts on $\chi(t)$ as a "low-pass filter" (see e.g. Bertero et al., 1995, Istratov and Vyvenko, 1999). This means that high frequencies in $\chi(t)$ are suppressed by convolution and show up damped in the response $\Delta Y(t)$. Hence, recoveries with large errors only at high frequencies tend to give relatively small prediction errors. Because of the low SNR, the pulse method yields a recovery of $\chi(t)$ with large errors both at high and low frequencies. Although the errors at high frequencies are damped in the prediction, errors at low frequencies are not. Hence, the large recovery error results in a large prediction error. On the other hand, because of the good SNR for the step response, the step method gives a relatively good recovery of $\chi(t)$ at low frequencies, with large errors concentrated at high frequencies. As a result, the large recovery error results only in a small prediction error. This suppression of high-frequency errors might also explain why the prediction error for

the step method remains unchanged when recovering the response function from a single response instead of the ensemble average. By comparing the recovery of $\chi(t)$ by the step method in Fig. 2.9a$_1$ and a$_2$, one sees that the main difference is indeed at high frequencies (the recovery in Fig. 2.9a$_2$ is quite "noisy" but follows the long term trend). This is because the noise amplification has a larger effect on the recovery from the single response due to its larger noise level. But since low frequencies are well recovered in both cases, the resulting prediction errors are almost the same.

Overall, the analysis of Fig. 2.9 suggests two main conclusions. First, as expected, the prediction error gives indeed an indication of the quality of the recovery, since good recoveries result in good predictions. But care should be taken when judging the recovery only from the prediction error, because a good prediction does not necessarily imply a good recovery: Due to the ill-posedness, eq. (2.1) might damp large high-frequency recovery errors so that they do not show up in the prediction. Nevertheless, from a good prediction error one can still infer a good recovery at low frequencies, because at these frequencies large recovery errors result in large prediction errors. Since regularization filtering leaves only low-frequency terms in the recovery, the RFI method shows in Fig. 2.9 small prediction errors associated to small recovery errors.

Second, by taking only a single response – and not the ensemble average – the full RFI algorithm gives on average smaller recovery and prediction errors than the pulse and step methods when comparing results obtained from the same experiment.

But the results above cover only the case where the full RFI algorithm is employed. In the following, I analyze also the case where monotonicity is not taken into account. For this purpose, I repeated in all detail the exercise that led to Fig. 2.9 but did not apply the additional noise level adjustment to enforce monotonicity of the response function. Fig. 2.10a shows the results for the recovery error. Once more, the RFI method gives smaller values than the step and pulse methods when comparing the recovery from the same responses. In addition, now the recovery for the RFI method using the step experiment even improved in comparison to Fig. 2.9b$_2$. The reason may be related to the numerical check for monotonicity: Depending on the tolerance value that is used to judge whether the recovered response function is monotonic, the additional adjustment might actually overestimate the noise level, leading to slightly worse results.

Yet, the improvement brought by the additional noise level adjustment is clear when looking at the recovery error for the 1% experiment. Compared to Fig. 2.9b$_2$, the average error increases substantially, and the spread is much larger (see inset for the whole value). As explained in section 2.4.4, this deterioration results from cases where the noise in the response is such that the spectral similarity assumption does not hold. Since here the noise estimate resulting from this assumption is not further improved by the monotonicity check, the result is a poor recovery (see subfigure (b) for an example). But because the large errors are mostly at high frequencies, even poor recoveries are still sufficiently good for predictions, as shown by the small mean prediction error in subfigure (c) (see "RFI$_{1\%}$"). Therefore, in contrast to the case where monotonicity is taken into account, here some small prediction errors are associated to large recovery errors.

Nevertheless, I find that although extreme, such poor recoveries are not frequent. In fact, extreme cases with recovery error $\varepsilon^r > 1$ account for 6.5% of the recoveries. This suggests that the large deterioration in the mean and spread of the recovery error in subfigure (a) is not a result of overall poor recoveries, but of only few extreme cases. To check this hypothesis, I plot in subfigure (d) the mean and standard deviation excluding these cases from the calculations. Indeed the result is much better, showing values comparable to the case where monotonicity is taken into account (compare "$\text{RFI}_{1\%}$" in Fig. 2.10d and Fig. 2.9b$_2$). Overall, this result indicates that at least for models of this type – where in the perturbation experiment the spectral distribution of noise does not change drastically compared to the control simulation – although monotonicity plays a role in avoiding large recovery errors, statistically most recoveries are still relatively good even without this additional improvement.

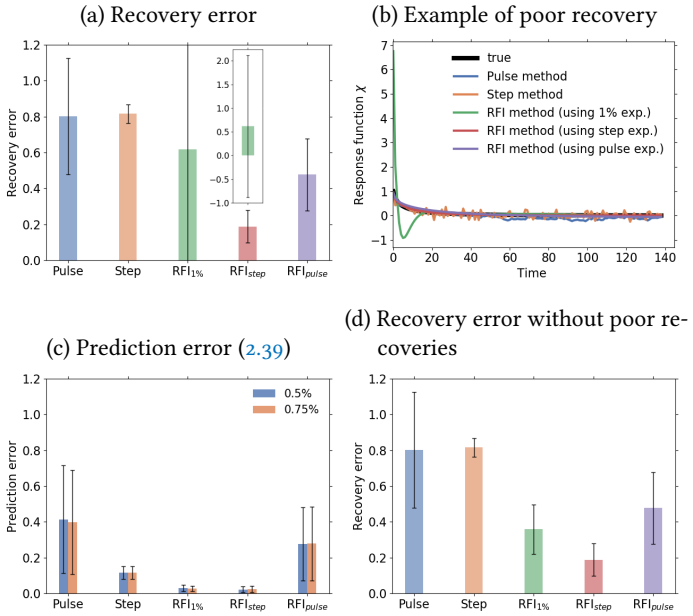


Figure 2.10 : Quality of response function recovery by our RFI method excluding step 6 in Fig. 2.2 in comparison to the pulse and step method. Response function is recovered taking the individual response for each ensemble member. Subscripts at "RFI" indicate the experiment from which the response function was recovered with the RFI method. (a) Statistics of the recovery error; (b) example of poor recovery with the RFI algorithm; (c) statistics of the prediction error (2.39); (d) statistics of the recovery error excluding for the $\text{RFI}_{1\%}$ the 6.5% of the recoveries with recovery error greater than 1. Once again, the RFI method gives better recovery and prediction errors than the Pulse and Step methods for the same responses. Without accounting for monotonicity the variability in the quality of the recoveries from the 1% experiment increases substantially, but poor recoveries are obtained only in few cases.

2.6 SUMMARY AND DISCUSSION

Existent methods to identify linear response functions from data require tailored perturbation experiments. Here, I developed a method to identify linear response functions from data using only information from an arbitrary perturbation experiment and a control simulation. The RFI method adresses the ill-posedness inherent to the identification problem by applying Tikhonov-Phillips regularization. The regularization parameter is computed by the discrepancy method, which involves the estimation of the noise level. For this purpose, I take advantage of information given by a spectral analysis of the perturbation experiment and by the control simulation. Assuming that the Picard condition holds, I estimate from the perturbation experiment the high-frequency components of the noise. Then, assuming that the spectral distribution of noise is approximately the same for the perturbed and control simulations (spectral similarity assumption), I estimate from the control simulation the low-frequency components of the noise. The obtained noise level estimate can be further adjusted if the linear response function is known to be monotonic. The robustness of the method in the presence of noise and nonlinearity was demonstrated in section 2.4.

As discussed in section 2.5, the developed method to identify linear response functions is an alternative approach to existent methods in the literature, which require special perturbation experiments and often give results with large errors caused by noise. In contrast, the RFI method accounts for the noise and can be directly applied to data from any type of perturbation experiment once also a control simulation is given. Because it accounts for the noise, its results have shown in the cases analyzed here a higher quality compared to results from previous methods when applied to the same data from a toy model. And because it can identify response functions from any type of perturbation experiment, the method is particularly suitable for application to data from model intercomparison exercises such as C^{4}MIP.

The main novelty of the method is the estimation of the noise level (steps 1-3 of Fig. 2.2), which is known to be critical for the application of regularization theory. When solving a problem by regularization, the most crucial step is the computation of the regularization parameter. To compute this parameter in a way that the solution converges to the "true" solution for decreasing noise, methods need to account for the noise level (Bakushinskii, 1984, Engl et al., 1996). But in practical applications the noise level is rarely known. Therefore, methods to obtain good estimates are needed. The noise estimation here consists essentially of two steps: First, estimating the high-frequency components from data and then the low-frequency components from the control experiment. The second step is completely novel, while the main idea behind the first step was already brought up in earlier studies (e.g. Hansen, 1990) and has recently been further developed by methods to compute the "Picard parameter" (Taroudaki and O'Leary, 2015, Levin and Meltzer, 2017). The "Picard parameter" is computed as the index for which the components $|\mathbf{u}_i \bullet \boldsymbol{\Delta}\mathbf{Y}|$ start to level off so that typically from this index onwards the data can be interpreted as noise. Although this levelling off is indeed often observed in artificial examples, the determination of noise components in this way depends on the spectral characteristics of the "true" data and of

the noise itself. For instance, if the noise components in the Picard plot decay smoothly as in Fig. 2.8a$_2$, detecting this parameter is not straightforward. In the RFI method, the interest lies in obtaining not all data components dominated by noise, but only enough components to obtain the overall level of the high-frequency noise. For this purpose, I define instead of the "Picard parameter" the more conservative index i_{max}, above which the singular values are zero and by the Picard condition also the "true" data components must be zero. In this way, I unambiguously identify data components that contain only noise (see eq. (2.27)). These components give the high-frequency level so that in the second step also the remaining low-frequency noise components can be estimated from the control experiment (step 3 of Fig. 2.2).

Because the noise level estimation here is not particularly related to the problem of identifying response functions, it can in principle be applied to solve also other types of linear ill-posed problems (see e.g. Engl et al., 1996). In general, all one needs for the application is:

1. A problem of the type $\mathbf{y} = A\mathbf{x} + \boldsymbol{\eta}$.
2. Data from a situation similar to the control experiment where $A\mathbf{x} = \mathbf{0}$, i.e. $\mathbf{y}_{ctrl} = \boldsymbol{\eta}_{ctrl}$.
3. The singular values of A decaying to values sufficiently close to zero to obtain i_{max}.

Then, as long as both the Picard condition and the spectral similarity assumption hold, the method gives a good noise estimate by which the regularization parameter can be determined.

While the Picard condition is necessary for a solution to be recoverable from an ill-posed problem, the validity of the spectral similarity assumption is less clear. Although no argument has been completely worked out in this direction, an intuitive explanation for this assumption can be thought as follows. Since here the interest lies on identifying linear response functions, the perturbation to the system must be sufficiently weak so that the response can be considered linear. If the noise in the control simulation depends on the perturbation, a sufficiently weak perturbation will modify its characteristics only slightly. The RFI method accounts partially for this change by adjusting the overall level by which the noise increases. Nevertheless, it assumes that since the characteristics of the noise change only slightly, then the spectral components of the noise in the perturbed simulation can be thought as having the same relative contributions as those in the control simulation. When in addition the response function is known to be monotonic, the estimate of the noise can be further improved (step 6 of Fig. 2.2), this time by adjusting the relative contribution of the spectral components: Since the high-frequency region is known from the spectral analysis of the response, then the components of the noise are adjusted in the low-frequency region; this is done iteratively until the resulting response function gets monotonic. Such additional adjustment has demonstrated to give good results in the applications in this and in the next chapters for the special case where the response function can be considered monotonic.

Although it is assumed that $\chi(t)$ is given by (2.12), this is not essential for the method. In principle, any functional form can be assumed for $\chi(t)$, or even none – in which case one would recover $\chi(t)$ pointwise. But compared to the simpler pointwise recovery of $\chi(t)$, assuming (2.12) has some advantages. The most obvious is that in contrast to the pointwise approach, with (2.12) both $\chi(t)$ and the spectrum can be recovered together. If $\chi(t)$ is recovered pointwise, the spectrum has to be derived in a second step from $\chi(t)$, which is also an ill-posed problem (Istratov and Vyvenko, 1999). Further, the description (2.12) restricts the function space for the recovered $\chi(t)$, forcing $\lim_{t\to\infty}\chi(t) = 0$ as is expected, which greatly simplifies the problem compared to the case where $\chi(t)$ can assume any form. On the other hand, the ansatz (2.12) is a restriction to the class of responses with τ being real. The most general ansatz would allow complex values for τ. Hence by (2.12) oscillatory contributions to $\chi(t)$ are excluded.

In addition to the application to identify climate-carbon cycle sensitivities aimed in this dissertation, the method developed in this chapter can be applied to investigate different aspects of the carbon cycle and climate. As a result of accounting for the memory of the system, the linear response function gives information on the strength of the response at all internal time scales covered by the times series underlying its recovery and probably even slightly longer (see next chapter). Since different processes have different characteristic time scales, a comparison of linear response functions from different models might help clarifying model differences by indicating for different models which time scales and thereby which processes are globally or regionally relevant. Using this method, in principle one can even compare the spectra of time scales from models to those from observations (e.g. Forney and Rothman, 2012b). The method presented here may also be applied to analyze changes in age and transit time distributions of carbon in different models, since these distributions can be derived directly from linear response functions (Thompson and Randerson, 1999). In all of the mentioned examples (see also section 2.1), the generality of the RFI method allows for the derivation of the appropriate linear response functions for any model by taking only data from an arbitrary perturbation experiment and a control simulation. Such generality opens the possibility of combining the linear response function framework, which has been gaining increasing attention due its wide applicability in climate science, with model intercomparison studies, hopefully leading to a deeper understanding of critical differences encountered across models.

3

LINEAR RESPONSE FUNCTIONS AS GENERALIZATION OF LAND CARBON SENSITIVITIES IN THE MPI-ESM

3.1 INTRODUCTION

In the last chapter it was seen that identifying the linear response function from data is not a trivial problem. Since this is needed for the application of the α-β-γ framework intended in dissertation, a method was developed for this purpose. The RFI method was tested in several applications to data taken from experiments performed with a toy model. Nevertheless, those tests represented an idealized situation, where the linear response function was already know, and conditions of noise and nonlinearities were controlled. But if one hopes to employ the linear response framework to study properties of the response of the global carbon cycle, one must be able to identify these functions as well from complex Earth System Models.

The purpose of this chapter is to show that by the RFI method one can indeed identify these functions from such a model. For this application, I take data from standard C^4MIP 1%-experiments performed with the Max Planck Institute for Meteorology Earth System Model (MPI-ESM, see Appendix A.1; for a description of the experiments see Table 2.1). In particular, I identify the linear response functions that generalize Friedlingstein's γ- and β-sensitivities for the land carbon, which will be referred to as land carbon *generalized sensitivities*. Since in this application the "true" response functions are not known, the robustness of the identified response functions will be demonstrated by checking their ability to predict the response of the MPI-ESM in experiments not used for the identification.

In the present chapter it will be also demonstrated that when these functions are derived with high quality, they provide useful information about the internal spectrum of time scales of the carbon cycle. To be able to obtain a recovery with high quality, different approaches are investigated that can be employed in conjunction with the RFI method to improve the recovery of the response functions. These approaches will be relevant here for recovering the spectrum of time scales and also in chapter4 to improve the accuracy of the description of the dynamics of the carbon cycle by the generalized α-β-γ framework.

As discussed in chapter 2, the presence of nonlinearities in the data can severely deteriorate the recovery of the response function. To make sure that the data from which the response function is recovered contain no strong nonlinearities, one must be able to estimate the linear regime of the response. Because the response functions will be derived from 1% experiments, in this chapter I introduce a technique to estimate with the aid of

additional simulations the linear regime in this type of experiment. By this technique the linear regime of the response of land carbon to changes in CO_2 and climate for the MPI-ESM will be estimated. Such estimate will give insight into what the linear regime might as well be in other models, for which additional simulations are not available. This idea will be employed to derive several linear response functions for different models in chapter 4.

The observables taken for the identification are the ones relevant for the quantification of the land carbon sensitivities γ and β, respectively:

(a) The change in global land carbon in response to changes in global land temperature;

(b) The change in global land carbon in response to changes in atmospheric CO_2.

Global land carbon is computed as the sum of the total land carbon over all grid cells of the model. Global land temperature is calculated as the mean near-surface air temperature over land at 2 m height. The changes are computed as $\Delta Y(t) = Y(t) - Y_{PI}$, with Y_{PI} being the mean value of observable Y from a control simulation at pre-industrial conditions. Since the main interest lies in long-term variations, annual mean data are used.

As demonstrated in Appendix A.4, the generalization of the β-sensitivity can be shown to be monotonic. Therefore, in the following I will derive it employing the additional noise level adjustment in the RFI algorithm (step 6 of Fig. 2.2). Since the generalization of the γ-sensitivity is not known to be monotonic, for this sensitivity the RFI algorithm will be applied without the additional adjustment.

The outline of the present chapter is the following. In the next section I introduce the technique to estimate the linear regime of the response in "percent" experiments. In section 3.3 I identify the generalization of the γ-sensitivity. In section 3.4 I identify the generalization of the β-sensitivity. In section 3.5 I present an example of a relatively detailed spectrum of time scales obtained for the response of land carbon to Net Primary Productivity, and how one can investigate the reasons for the observed structure of the spectrum. In section 3.6 the results are summarized and discussed.

3.2 ESTIMATING THE LINEAR REGIME IN "PERCENT" EXPERIMENTS

As described in chapter 2, the recovery of response functions is cursed by the presence of noise and nonlinearities. The RFI method is designed to cope with the former. In the present section I present a technique to cope with the latter, but at the expense of performing additional response experiments. This technique will serve as a complement to the RFI method in the application to the land carbon cycle in the following sections and in the application to the global carbon cycle in chapter 4. By these additional experiments I will determine the range of forcing strengths for which the response can be considered to be linear – an analysis in general not possible using only the control and perturbation experiments on which the RFI method is based.

I introduce this technique by the example of toy model simulations. Taking first a purely linear situation ($a = 0$) I show in Fig. 3.1a the prediction error (2.39) when using the response function obtained from a 1% simulation to predict the response from two other %-simulations with smaller growth rate. More precisely, performing a sequence of 1% experiments for increasingly longer simulation periods, I calculated for each experiment the response function and used it to predict the response for a 0.5% and 0.75% experiment covering the same simulation period. Then I plotted in Fig. 3.1a the prediction error against the final forcing strength of the 1% experiment. As a result, it is seen that with increasing final forcing strength the prediction error decreases. This happens because in this linear case the SNR is increasing with increasing simulation period, i.e. with increasing final forcing strength, so that the recovery of the response function continuously improves.

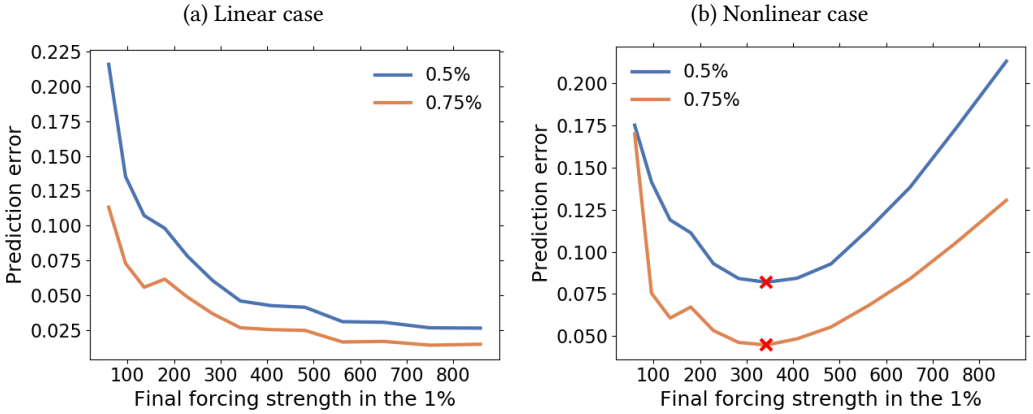


Figure 3.1 : Toy model example for the identification of the linear regime by using additional experiments. Shown is the prediction error (2.39) for the response of 0.5% and 0.75% experiments as obtained from the response function calculated by the RFI method from 1% experiments. The prediction errors are plotted against the final forcing strengths of a sequence of 1% experiments with increasing time series length. The crosses at the minima indicate the final forcing strength for which the response function is optimally recovered (see text). Subfigure (a) shows the behaviour for the fully linear toy model ($a = 0$) and subfigure (b) the behaviour in the presence of nonlinear contributions to the response ($a = 5 \times 10^{-5}$). For the purpose of demonstrating more clearly the increase in the prediction error for a decrease in the forcing strength, I include in the plot cases where the forcing strength is extremely small, corresponding to very small time series lengths. To deal with such cases, I set for a number of data points $N < 30$ the number of time scales $M = N$. For such small number of time scales, usually no plateau in the singular values spectrum is found (step 2 of Fig. 2.2). Therefore, for these special cases I also modify the algorithm to interpret the two smallest singular values as a plateau, since their small magnitude makes them have a similar effect to those singular values belonging to the plateau itself. In addition, to illustrate the most general case where $\chi(t)$ is not known to be monotonic, I exclude here the monotonicity check (step 6 of Fig. 2.2).

Calculating the prediction error only for experiments with smaller growth rate gets important in the next case where nonlinearities are considered (Fig. 3.1b). This plot was obtained by the same procedure except that I took for the nonlinearity parameter a value $a > 0$. As seen, in this nonlinear case the prediction error is first improving but deteriorating afterwards. For small forcing, nonlinearities are small and therefore the prediction error behaves as in the linear case, i.e. it decreases with final forcing strength. But when forcing strengths get larger, nonlinearities start to contribute substantially to the response, thereby causing a deterioration of the recovery of the response function and consequently the prediction error once more increases. This increase of the prediction error can be unambiguously traced back to the presence of nonlinearities in the 1% simulation because the prediction error was calculated only for experiments with smaller growth rate, i.e. smaller forcing strength throughout the whole simulation. Therefore, nonlinearities contribute already substantially to the response in the 1% simulation before they get relevant in the other experiments. Accordingly, with this type of experiment setup we can be sure that the increase in the prediction error comes solely from the deterioration of the recovery of the response function and not from nonlinearities in the additional experiments used for calculating the prediction error.

Obviously, for forcing strengths smaller than at the minimum, nonlinearities do not hinder the recovery of the response function so that one can consider this to be the regime of linear system behaviour. In view of the trade-off between noise and nonlinearities, for the 1% experiment the response function is thus optimally recovered when taking as final forcing strength the value at the minimum of the prediction error curve. Similarly, if the error curve has no minimum (as in the linear case shown in Fig. 3.1a) the optimal recovery is obtained from the experiment with the largest forcing strength.

With the presentation of this additional technique to identify the linear regime we are ready for the application to the MPI-ESM in the next sections.

3.3 GENERALIZED SENSITIVITY χ_γ

In this section, I identify from MPI-ESM simulations the linear response function χ_γ (generalization of the γ-sensitivity), defined by

$$\Delta C^{rad}(t) = \int_0^t \chi_\gamma(t-s)\Delta T(s)ds + \eta(t), \tag{3.1}$$

where now the response is $\Delta Y(t) := \Delta C^{rad}(t)$ and the forcing is $\Delta f(t) := \Delta T(t)$, with $\Delta C^{rad}(t)$ being the change in global land carbon obtained in the "rad" experiment and ΔT the change in global land temperature.

Estimating the linear regime

As a first step in obtaining a proper approximation of χ_γ, I investigate what maximum forcing strength can be used to assure that the recovered χ_γ is not spoiled by the presence of nonlinearities. Using the technique introduced in section 3.2, I show in Fig. 3.2 the prediction error (2.39) for $\chi_\gamma(t)$ recovered from the 1% rad experiment as a function of the final forcing strength in the 1% rad experiment. There is no clear minimum so that for the data available the recovery of $\chi_\gamma(t)$ seems not to be limited by nonlinearities. For optimal recovery I thus take the the full time series, i.e. the maximum final forcing strength.

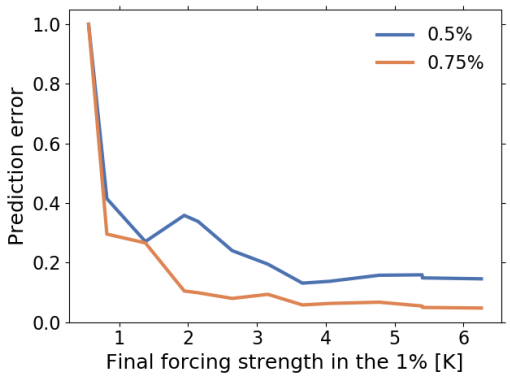


Figure 3.2 : Prediction error (2.39) for the 0.5% and 0.75% rad experiments using $\chi_\gamma(t)$ recovered from the 1% rad experiment. The error is shown as function of the final forcing strength used for the recovery of $\chi_\gamma(t)$. No clear minimum is found so that the recovery seems not to be limited by nonlinearities.

χ_γ and the quality of its recovery

The quality of the recovery can in principle be improved by taking an experiment with better SNR. To investigate if the recovery from the 1% rad experiment can be further improved, I applied the RFI algorithm also to recover χ_γ from the $2\times CO_2$ rad experiment. I chose the $2\times CO_2$ rad experiment because as shown in Fig. 3.3 it has smaller forcing strengths than the maximum forcing strength for the 1% rad experiment – therefore nonlinearities should also not limit the recovery – but is expected to carry useful information over a larger range of the response spectrum. This expectation can be justified as follows (MacMartin and Kravitz, 2016). Taking the Laplace transform of (3.1) gives

$$\Delta\widetilde{C}^{rad}(p) = \widetilde{\chi}_\gamma(p)\Delta\widetilde{T}(p) + \widetilde{\eta}(p), \tag{3.2}$$

where the tilde denotes Laplace transformed functions. From Fig. 3.3 it is seen that for the 1% rad experiment the temperature ΔT behaves approximately as a linear function, which gives a Laplace transform $\Delta\widetilde{T}(p)$ proportional to $1/p^2$. For the $2\times CO_2$ rad experiment, the temperature behaves approximately as a step function (ignoring the transient in the first

20 years), which gives a Laplace transform $\Delta\widetilde{T}(p)$ proportional to $1/p$. This means that for the same χ_γ, the first term on the right-hand side of (3.2) – the "clean" response – decays to zero faster for the 1% rad experiment than for the $2\times CO_2$ rad experiment. Hence, assuming the same noise η for both cases, the response from the $2\times CO_2$ rad experiment gets buried in the noise only at larger p, meaning that this experiment carries useful information until higher rate values p than the response from the 1% rad experiment.

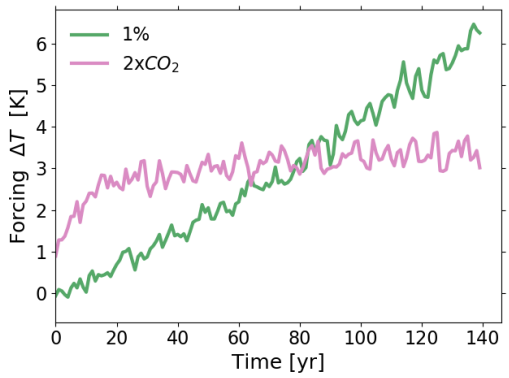
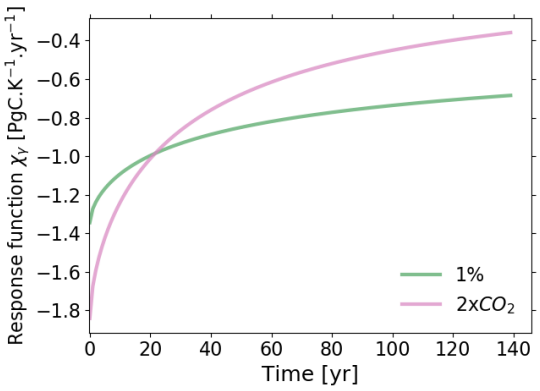


Figure 3.3 : Forcing temperature $\Delta T(t)$ for 1% and $2\times CO_2$ rad experiments.

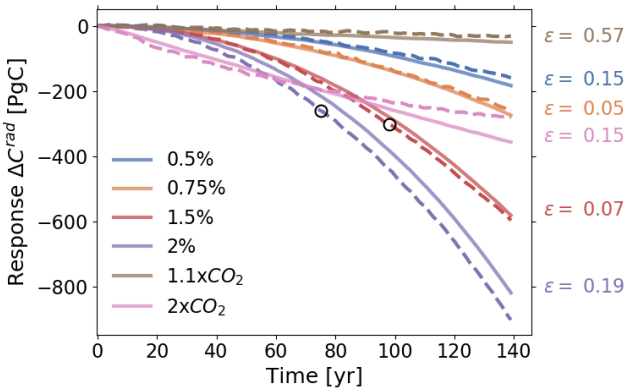
The response function $\chi_\gamma(t)$ recovered from the two types of experiments is shown in Fig. 3.4a. As expected, the different experiments indeed result in different recoveries. Because we know from the analysis of Fig. 3.2 that nonlinearities do not limit the recovery, the difference between the two response functions results probably from the difference in the quality of the data from the two types of experiments. To compare the robustness of each recovery, I analyze how well they predict the response from other experiments (Fig. 3.4b and c). If the response function is correctly recovered, it should be able to predict not only experiments with smaller but also experiments with higher forcing rate. Therefore, I include in the analysis also 1.5% and 2% rad experiments. To exclude errors that may be caused by nonlinearity, I take as a conservative estimate of the linear regime forcing strengths smaller than the final forcing strength at the end of the 1% rad experiment (which is approximately the maximum forcing strength; see temperature value at t = 140 years for the 1% rad experiment in Fig. 3.3). I take these values as an estimate of the linear regime because the 1% rad experiment has only 140 years so that no estimate for higher forcing strengths is available. Hence, for the 1.5% and 2% rad experiments the responses are expected to be reasonably predictable at least until the values marked with circles, where their respective forcing strengths reach this maximum forcing strength. All other experiments have forcing strengths smaller or equal to this maximum forcing strength so that they should be predictable for the whole time series.

Fig. 3.4b shows the quality of the prediction using $\chi_\gamma(t)$ recovered from the 1% rad experiment. Visually, model response and prediction seem to have a comparable quality of agreement across the $1.1\times CO_2$, 0.5%, 0.75% and 1.5% rad experiments, while for the 2%

(a) χ_γ derived from 1% and 2×CO_2 rad



(b) Prediction from χ_γ derived from 1% rad



(c) Prediction from χ_γ derived from 2×CO_2 rad

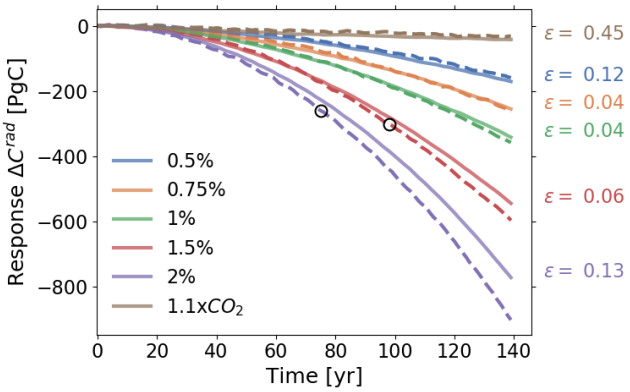


Figure 3.4 : χ_γ derived from 1% and 2×CO_2 rad experiments and prediction of model responses using these χ_γ (continuous lines are predictions and dashed lines are responses from the MPI-ESM). Circles indicate the maximum value for which 1.5% and 2% responses are predictable according to the estimate of the linear regime (see text). At the right of subfigures (b) and (c) the prediction error (see eq. (2.39)) is indicated for the different experiments, calculated for the 1.5% and 2% rad experiments by considering only values preceding the circles.

and $2\times CO_2$ rad experiments there are larger discrepancies. For a quantitative analysis, I compute for the estimated linear regime the prediction error (2.39) for each experiment (right side of the plot). It is seen that the error varies from less than 10% for the 0.75% and 1.5% rad experiments to values between 10-20% for the 0.5%, 2% and $2\times CO_2$ rad experiments, and a significantly larger value of 57% for the $1.1\times CO_2$ rad experiment. To better understand these differences, it is important to note that as long as nonlinearities are small, experiments with higher forcing strength are expected to have smaller prediction error because they have higher SNR. This can be made plausible by considering that a perfectly recovered response function predicts a "clean" linear response (infinite SNR) with zero error, whereas the same response function can predict a noisy response only with some finite error. Therefore, if $\chi_\gamma(t)$ is well recovered, we expect large prediction errors for experiments with small forcing strengths such as the $1.1\times CO_2$ rad – which is indeed the case –, but small errors for experiments with large forcing strengths but still well within the linear regime such as the $2\times CO_2$ rad (compare the forcing strengths for the $2\times CO_2$ rad experiment and the maximum forcing strength for the 1% rad experiment in Fig. 3.3). Since contrarily to the expectation the prediction error is not particularly small for the $2\times CO_2$ rad experiment, probably the recovery of $\chi_\gamma(t)$ derived from the 1% rad experiment is not completely accurate and may still be further improved. As suggested above, such improvement may be achieved by taking data with better quality from the $2\times CO_2$ rad experiment.

Fig. 3.4c shows the quality of the prediction using $\chi_\gamma(t)$ recovered from the $2\times CO_2$ rad experiment. As expected, results indicate an improvement in the recovery (compare to subfigure (b)). The prediction error decreases for all experiments present in both plots. In addition, it also decreases if we compare the prediction of the 1% rad response in subfigure (c) with that of the $2\times CO_2$ rad response in subfigure (b).

This subsection therefore suggests two main conclusions. First, for $\chi_\gamma(t)$ the response seems to be approximately linear until temperature perturbations of at least 6 K. Second, the overall improvement of the prediction in Fig. 3.4c compared to Fig. 3.4b confirms the expectation from the Laplace transform analysis that indeed the step-like $2\times CO_2$ rad experiment carries more information on the response function than the 1% rad experiment. This suggests that step-like experiments may be more appropriate than the standard 1% experiment for the identification of linear response functions.

3.4 GENERALIZED SENSITIVITY χ_β

In this section, I identify the linear response function χ_β (generalized β-sensitivity), defined by

$$\Delta C^{bgc}(t) = \int_0^t \chi_\beta(t-s)\Delta c(s)ds + \eta(t), \tag{3.3}$$

where now the response is $\Delta Y(t) := \Delta C^{bgc}(t)$ and the forcing is $\Delta f(t) := \Delta c(t)$, with $\Delta C^{bgc}(t)$ being the change in global land carbon found in the "bgc" experiment and Δc the change in atmospheric CO_2.

Approaches to identify χ_β

Similarly to last section, I identify $\chi_\beta(t)$ by several approaches to find the one that gives results with best quality. Here, the identification is performed in three different ways:

1. Using CO_2 as forcing (see eq. (3.3));
2. Using the logarithm of CO_2 as forcing:

$$\Delta C^{bgc}(t) = \int_0^t \chi_\beta^{\ln}(t-s) c_{PI} \ln\left(\frac{c(s)}{c_{PI}}\right) ds + \eta(t), \tag{3.4}$$

   where the forcing is $\Delta f(t) := c_{PI} \ln\left(\frac{c(t)}{c_{PI}}\right)$, with c_{PI} being the pre-industrial value for atmospheric CO_2;
3. Using Net Primary Production (NPP) as forcing:

$$\Delta C^{bgc}(t) = \int_0^t \chi_{NPP}(t-s) \Delta NPP(c(s)) ds + \eta(t). \tag{3.5}$$

The first approach is the same used throughout the paper: $\chi_\beta(t)$ is identified using the $\Delta c(t)$ forcing from eq. (3.3).

For the second and third approaches, I take advantage of the fact that in the "bgc" setup, perturbations in atmospheric CO_2 affect land carbon only via changes in photosyntetic productivity. Therefore, I use the CO_2 forcing only indirectly via its relationship to NPP. By doing this, the hope is to account for some of the nonlinear contributions to the response that arise from this relationship. The advantage of accounting for these nonlinear contributions is that one can recover the response function from experiments with perturbation strengths larger than those possible when not accounting for nonlinear contributions (as in the first approach). Experiments with larger perturbation strengths give responses with higher SNR, which makes it possible to recover the response function with better quality.

In the second approach (see eq. (3.4)), I employ as forcing an explicit logarithmic expression describing the relationship between CO_2 and NPP (see e.g. Alexandrov et al., 2003). Such formula has the advantage that the expansion of the forcing in c gives

$$\Delta C^{bgc}(t) = \int_0^t \chi_\beta^{\ln}(t-s) \Delta c(s) ds + \mathcal{O}((\Delta c)^2) + \eta(t). \tag{3.6}$$

Taking Δc sufficiently small and comparing the result to (3.3) thus yields

$$\chi_\beta(t) = \chi_\beta^{\ln}(t). \tag{3.7}$$

Accordingly, the response function $\chi_\beta^{\ln}(t)$ from formula (3.4) gives as well the desired $\chi_\beta(t)$.

In the third approach, I take directly the response to NPP (see eq. (3.5)). Expanding the forcing $\Delta NPP(c)$ in c gives

$$\Delta C^{bgc}(t) = \int_0^t \chi_{NPP}(t-s)\frac{\partial NPP}{\partial c}\Big|_{c=c_{PI}} \Delta c(s)ds + \mathcal{O}((\Delta c)^2) + \eta(t). \tag{3.8}$$

Taking Δc sufficiently small and comparing the result to (3.3) yields

$$\chi_\beta(t) = \chi_{NPP}(t)\frac{\partial NPP}{\partial c}\Big|_{c=c_{PI}}. \tag{3.9}$$

Accordingly, by this approach I compute $\chi_\beta(t)$ in three steps: First, I identify the response function $\chi_{NPP}(t)$ using (3.5); second, I take the first derivative of NPP with respect to CO_2 at $c = c_{PI}$; third, I apply formula (3.9) to obtain $\chi_\beta(t)$ from $\chi_{NPP}(t)$.

Checking nonlinearities with the three approaches

Before analyzing the recovery of $\chi_\beta(t)$ employing the three approaches, one must verify that indeed these two additional approaches account for some of the nonlinearities in the response. If this is true, response formulas (3.4) and (3.5) should be able to predict the response to larger perturbation strengths than (3.3). To verify this expectation, in the following I compare the prediction error (2.39) by applying eqs. (3.3), (3.4) and (3.5).

Fig. 3.5a shows the prediction error for $\chi_\beta(t)$ recovered from the 1% bgc experiment with the first approach (eq. (3.3)). The prediction is also computed via eq. (3.3). The clear minima indicate the presence of strong nonlinearities for forcing strengths above around 100 ppm (94 ppm for the 0.5% and 133 ppm for the 0.75%). Therefore, in contrast to the case of χ_γ discussed in the last section, we see here that indeed one has to cope with the additional difficulty of nonlinearities.

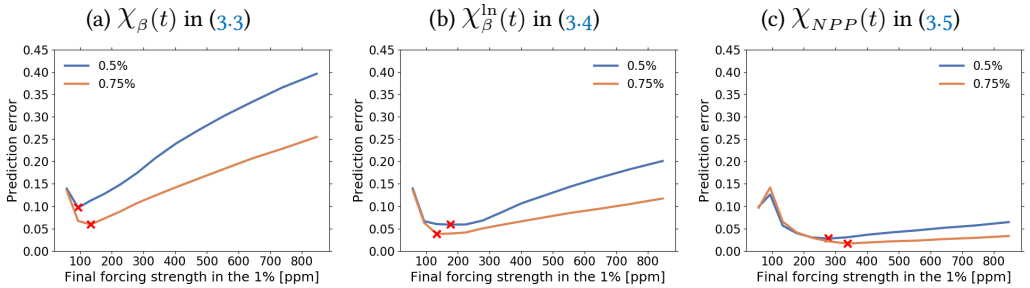


Figure 3.5 : Prediction error (2.39) for the 0.5% and 0.75% bgc experiments obtained when using $\chi_\beta(t)$, $\chi_\beta^{\ln}(t)$ and $\chi_{NPP}(t)$ obtained from the 1% bgc experiment to predict the response. The error is shown as a function of the CO_2 final forcing strength.

Fig. 3.5b shows the prediction error when using $\chi_\beta^{\ln}(t)$ recovered from the 1% bgc experiment with the second approach (eq. (3.4)). To check how well nonlinearities are accounted for by taking the logarithmic forcing, the prediction is as well computed via eq. (3.4). Compared to subfigure (a), we see a slight improvement in the results: The minima have smaller prediction errors and the prediction errors increase at a slower rate for increasing final forcing strength. This indicates that indeed using the logarithm of CO_2 as forcing accounts for some of the nonlinearities in the response. Accordingly, one can make predictions with smaller error for larger forcing strengths using (3.4) instead of (3.3).

Fig. 3.5c shows the prediction error for $\chi_{NPP}(t)$ recovered via eq. (3.5) from the 1% bgc experiment (first step of the third approach). To check how well nonlinearities are accounted for by taking the NPP forcing, I employ as well eq. (3.5) for the prediction. Here, we see a substantial improvement in the results. The response is almost completely linear, with very "flat" minima. This indicates that indeed a large part of the nonlinearity encountered in the response of the land carbon to changes in CO_2 can be explained by the nonlinear relationship between NPP and CO_2. Accordingly, by employing (3.5) instead of (3.3) one can predict the response of the land carbon until forcing strengths as high as 800 ppm with an error smaller than 10%.

χ_β and the quality of its recovery

So far, I have only considered the ability of eqs. (3.3), (3.4) and (3.5) to predict the land carbon response. Now, I analyze how well the generalized sensitivity $\chi_\beta(t)$ can be identified by the three approaches. For the identification I took data from the 1% bgc experiment until the CO_2 forcing strength reaches the first minimum for each case in Fig. 3.5: Δc = 94 ppm for the first approach (30 years); Δc = 133 ppm for the second approach (40 years); and Δc = 279 ppm for the third approach (70 years). Since Δc = 279 ppm is approximately the forcing strength for the $2\times CO_2$ bgc experiment and results from last section suggested that this type of experiment may carry more information for the identification, I employ the third approach also taking the $2\times CO_2$ bgc experiment. For the present application where the recovery is limited by nonlinearities, taking the $2\times CO_2$ bgc experiment has the additional advantage that because its forcing strength has a constant value throughout the whole experiment, we can use the full time series (140 years). To compute the first derivative of NPP with respect to CO_2 (second step of the third approach), I fitted the function $NPP = NPP(c)$ to polynomials of order 4, 5 and 6, and then took the first derivative from the fits.

The results from the three approaches are shown in Fig. 3.6a. At short time scales there is an overall agreement among all recoveries with only small discrepancies. To be able to compare the results also for longer time scales, I extend the response functions recovered from the 1% bgc experiment – obtained from time series with 30, 40 and 70 years respectively for the first, second and third approaches – until 140 years (extensions are indicated by dotted lines). This can be done because with the RFI method I derive the response function

from the ansatz (2.12), which formally gives the values of the response function for all times. The result is that all response functions recovered from the 1% bgc experiment present relatively small discrepancies even at long time scales. Response functions derived from the $2\times CO_2$ bgc experiment with the third approach show a similar behaviour among themselves, but differ from the recoveries using the 1% bgc experiment. The reason for this difference will be investigated below.

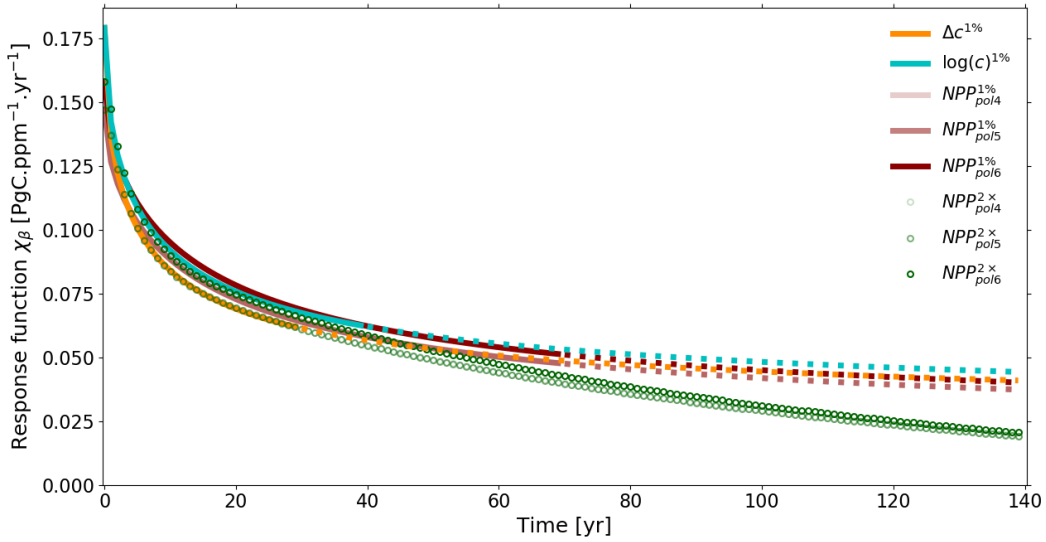
To quantitatively compare the quality of the recoveries, I plotted in Fig. 3.6b the prediction error (2.39). Since the response functions were derived using different time series lengths, for a fair comparison I compute the error only at the minimal time series length of 30 years (the time series length used for the first approach). Results show no large discrepancies among the different approaches. For the third approach, there seems to be a small advantage in using a polynomial of order 6 for the computation of the derivative.

But results from subfigure (b) reflect only the quality of the recoveries at short time scales, for which anyway no large discrepancies were encountered in subfigure (a). To evaluate the quality of the recovery also at long time scales, one must take the extended version of the response functions that were derived from shorter time series. Since the only substantial difference at long time scales in Fig. 3.6a is found between the response functions recovered from the 1% and $2\times CO_2$ bgc experiments, I take for this analysis exemplarily only the response functions recovered with the first approach (1% bgc experiment) and the third approach ($2\times CO_2$ bgc experiment). By choosing the response function recovered with the first approach, I evaluate for the worst case scenario (where only 30 years are used for the recovery) how reliable predictions are for time scales longer than the time series used for recovery. In contrast, by choosing the response function recovered with the third approach from the $2\times CO_2$ bgc experiment, I check whether the different values at long time scales are actually an improvement in the recovery. As mentioned above, such improvement is expected because this response function was recovered taking the full time series (140 years).

Following the same procedure as in the last section, in Fig. 3.7 I show the quality of the prediction for different experiments using the aforementioned recoveries of $\chi_\beta(t)$. Subfigure (a) shows the results for the predictions calculated via eq. (3.3) using $\chi_\beta(t)$ recovered with the first approach. I take as an estimate of the linear regime forcing strengths smaller than the forcing strength at the first minimum in Fig. 3.5a. The response values corresponding to this forcing strength are marked in Fig. 3.7a with circles. It is seen that although $\chi_\beta(t)$ was recovered using a time series of only 30 years, it predicts the $1.1\times CO_2$ bgc experiment with only 3% error over 140 years. Other experiments are predicted within the linear regime with error smaller than 10% with exception of the 2% bgc experiment, for which the error is around 14%. Since the $2\times CO_2$ has a constant forcing strength outside the linear regime already from the beginning, its prediction fails as expected for the whole time series.

In subfigure (a) we could only evaluate the quality of the long time scales of $\chi_\beta(t)$ by the prediction of the $1.1\times CO_2$ bgc experiment, because this is the only experiment which has forcing strengths within the linear regime over the whole time series. To check the

(a) Recovered $\chi_\beta(t)$



(b) Prediction error (2.39)

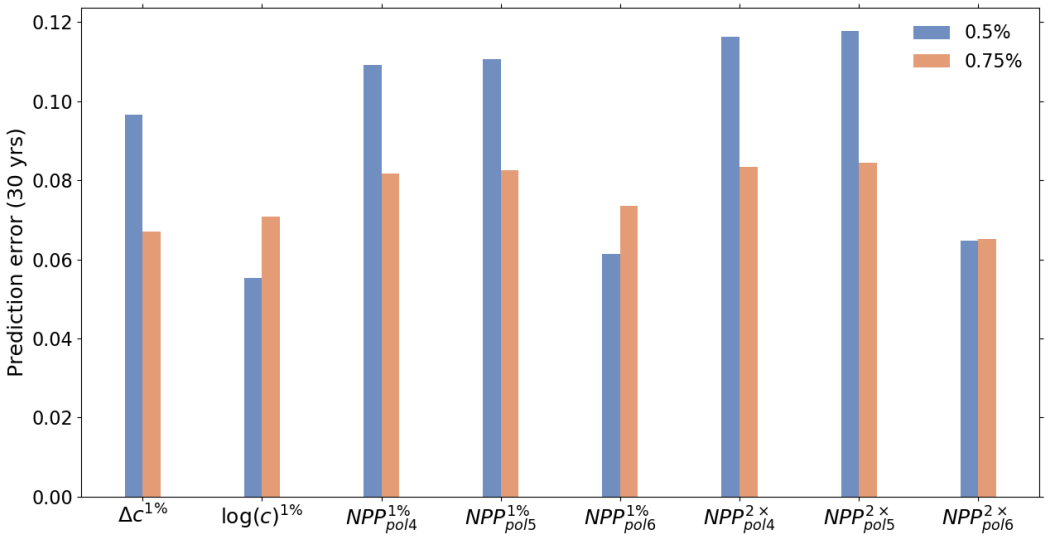
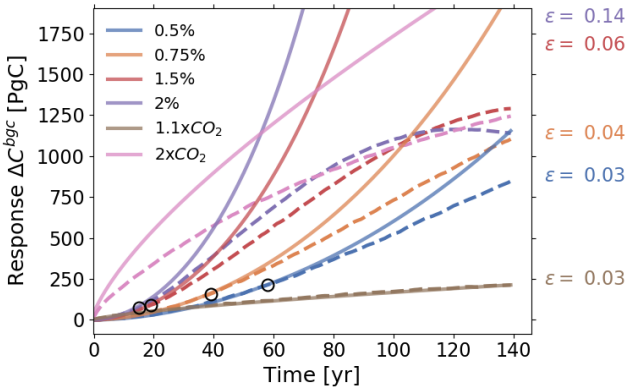
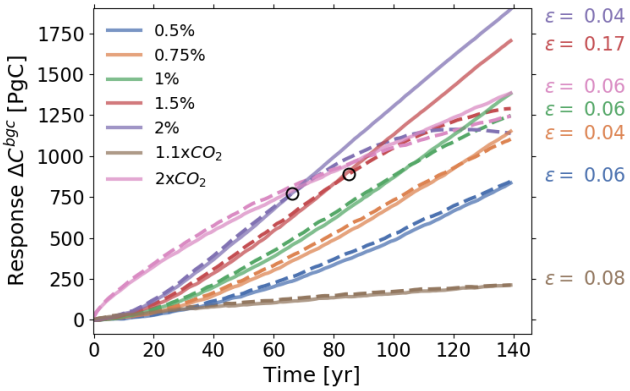


Figure 3.6 : Response function $\chi_\beta(t)$ derived by the three approaches (subfigure (a)) and the respective prediction errors for the first 30 years of the response (subfigure (b)). $\Delta c^{1\%}$: recovery with first approach from 1% bgc experiment; $\log(c)^{1\%}$: recovery with second approach from the 1% bgc experiment; $NPP^{1\%}_{polx}$: recovery with third approach from the 1% bgc experiment using for the derivative a polynomial fit of order x; $NPP^{2\times}_{polx}$: recovery with third approach from the $2\times CO_2$ bgc experiment using for the derivative a polynomial fit of order x. Continuous lines denote values of $\chi(t)$ that are within the time series length used for the recovery (30 years for $\Delta c^{1\%}$, 40 years for $\log(c)^{1\%}$ and 70 years for $NPP^{1\%}_{polx}$). Dotted lines denote extended parts of the response function, i.e. values not covered by the time series used for the recovery but obtained from the recovered spectrum using (2.12). Circles denote the response functions derived taking the full time series ($NPP^{2\times}_{polx}$). For more details see text.

(a) $\chi_\beta(t)$ from first approach using (3.3)



(b) $\chi_{NPP}(t)$ from first approach using (3.5)



(c) $\chi_{NPP}(t)$ from third approach using (3.5)

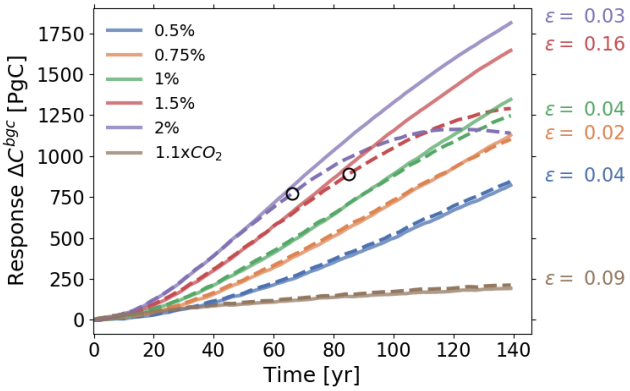


Figure 3.7 : Prediction of model responses employing response functions derived with the first approach from the 1% bgc experiment (derived from data with 30 years length but extended to 140 years) and with the third approach from the 2×CO_2 bgc experiment (derived from data with 140 years length). (a) Prediction by eq. (3.3) taking the response function $\chi_\beta(t)$ derived with the first approach. (b) Prediction by eq. (3.5) taking the response function $\chi_\beta(t)$ derived with the first approach and converted to $\chi_{NPP}(t)$ by eq. (3.9). (c) Prediction by eq. (3.5) taking the response function $\chi_{NPP}(t)$ derived in the first step of the third approach (see explanation after eq. (3.9)). Continuous lines are predictions and dashed lines are responses from the MPI-ESM. Circles indicate the maximum value for which responses are predictable according to our estimate of the linear regime (see text). The values printed to the right of the plots are the prediction errors (see eq. (2.39)) calculated for each experiment, considering when applicable only values preceding the circles.

ability of $\chi_\beta(t)$ to predict also other experiments at long time scales, I account for some of the nonlinearity in the response by taking NPP instead of CO_2 as forcing (see discussion of Fig. 3.5). Therefore, I perform the prediction by employing (3.5) instead of (3.3). Since for employing (3.5) one needs $\chi_{NPP}(t)$ and not $\chi_\beta(t)$, I take the $\chi_\beta(t)$ derived with the first approach and compute $\chi_{NPP}(t)$ from it via eq. (3.9). Because the conversion from $\chi_\beta(t)$ to $\chi_{NPP}(t)$ is a simple scaling, the time scales structure is maintained. Hence, we can evaluate the quality of the recovered $\chi_\beta(t)$ from the results given by the obtained $\chi_{NPP}(t)$. The prediction results computed via (3.5) with the obtained $\chi_{NPP}(t)$ are shown in subfigure (b). Because errors at the minima in Fig. 3.5c are not substantially smaller than those at the maximum final forcing strength, I take as an estimate of the linear regime all values smaller than the last value of NPP for the 1% bgc experiment (marked with circles in the responses). Once again it is seen that although $\chi_\beta(t)$ was recovered using a time series of 30 years, after conversion to $\chi_{NPP}(t)$ almost all experiments can be predicted with less than 10% error for the whole time series. The 1.5% and 2% bgc experiments are predicted with errors of 17% and 4% within the linear regime. Results from subfigures (a) and (b) therefore suggest that although nonlinearities do restrict the recovery from (3.3), taking experiments with forcing strengths within the linear regime for the recovery leads to reliable prediction results even for times reasonably longer than the length of the time series from which the response function was recovered.

Finally, I investigate whether the different values seen in Fig. 3.6a for the $\chi_\beta(t)$ recovered from the $2\times CO_2$ bgc experiment indeed reflect a better quality of recovery. Following the same reasoning that led to subfigure (b), since in the third approach $\chi_\beta(t)$ is obtained from a scaling of $\chi_{NPP}(t)$, I evaluate the quality of $\chi_\beta(t)$ from the results given by $\chi_{NPP}(t)$. Accordingly, in subfigure (c) I plot the prediction via (3.5) using the $\chi_{NPP}(t)$ recovered from the $2\times CO_2$ bgc experiment. As expected, the overall prediction indeed improves compared to subfigure (b). Individual prediction errors decrease for all experiments with exception of the $1.1\times CO_2$. Since the response functions used for subfigures (b) and (c) differ only at long time scales, this improvement suggests that indeed obtaining $\chi_\beta(t)$ from the $2\times CO_2$ bgc experiment gives a better recovery over these time scales. Further, because all recoveries are similar at short time scales (see Fig. 3.6a), overall this recovery shows the best quality.

3.5 SPECTRUM OF LAND CARBON TIME SCALES

A response function obtained with high quality not only results in more accurate predictions, but also provides valuable information about the internal dynamics of the system. For the case of $\chi_\beta(t)$, I find that the recovery with best quality gives a relatively detailed description of the spectrum of land carbon time scales (see eq. (2.12)). In this section, I analyze the robustness of this result and illustrate how possible explanations for it can be investigated. Since the best recovery was obtained from the response to NPP for the $2\times CO_2$ bgc experiment, in the following I study only this case.

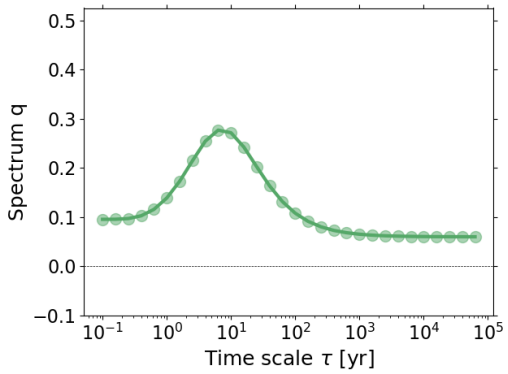
Obtaining the detailed spectrum and checking its robustness via a Gregory plot approach

To first understand how a better recovery improves the description of the spectrum, in Fig. 3.8a I plot the spectrum $q(\tau)$ of the response function used for Fig. 3.7b, i.e. $\chi_\beta(t)$ recovered with the first approach and converted to $\chi_{NPP}(t)$. Because the data used for recovery has a relatively low quality – the response function was recovered from the 1% bgc experiment taken for only 30 years –, regularization filters out most components of the spectrum in eq. (2.23). The result is the smooth spectrum shown in the figure which, albeit sufficiently good for making the predictions shown in Fig. 3.7b, is not very informative about the internal time scales structure. Instead, the spectrum of the response function $\chi_{NPP}(t)$ used for Fig. 3.7c has a more detailed structure (Fig. 3.8b). In this case, the higher quality of the data used for the identification ($2\times CO_2$ bgc experiment taken for 140 years) results in less filtering by regularization. The result is a spectrum with two peak time scales, around 4 and 100 years[1].

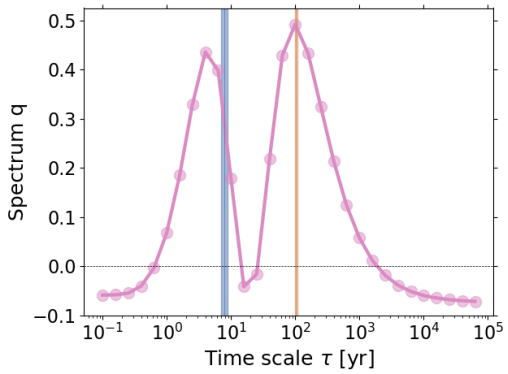
Some trust in this result may be gained by constructing a type of "Gregory plot" (Gregory et al., 2004) for the land carbon (subfigure (c)). The idea here is to try to identify from an independent method important time scales in the response so that they can be compared against the spectrum in subfigure (b). For this analysis, I plot the global land carbon against its first time derivative (this is the net land-atmosphere carbon flux). Using the $2\times CO_2$ bgc experiment where the forcing is constant, the first time derivative vanishes as the land carbon approaches a new equilibrium value. The rate at which the derivative changes can be associated to a time scale τ_i, which should show up with a large weight value q_i in the spectrum. Interestingly, the plot shows that the transient behaviour towards equilibrium develops approximately at two different rates: a higher rate from the starting point until around 3520 PgC, and a lower rate from 3520 PgC onwards. To determine these rates values, I fitted a linear function $dC/dt = a + bC^{bgc}$ for each of these two ranges of the land carbon. The obtained rate values b are shown in the legend. Then, from each rate value b I computed a time scale by $\tau = -1/b$. The computed time scales taking one standard deviation into account are shown by the two ranges highlighted in subfigure (b). As seen, while the shortest time-scale peak of the spectrum partially overlaps with the corresponding time-scale range from the Gregory plot, the longest time-scale peak is almost perfectly matched. This result suggests that indeed the recovered spectrum reflects internal characteristics of the global response of the land carbon cycle.

1 I ignore the negative values obtained for time scales smaller than 1 year (data time step) and larger than 140 years (time series length) because spectral contributions at time scales much longer or much shorter than the time scales covered by data cannot be correctly recovered. Yet, as demonstrated in Appendix B.1, their wrong recovery does not strongly affect the recovered response function so that they can be safely ignored. Other slightly negative values between the two peaks are probably small recovery errors that inevitably appear as a consequence of ill-posedness and the filtering by regularization (such slightly negative values have also been observed in recovered spectra shown in Fig. B.2 in Appendix B.2).

(a) First approach (1% bgc experiment)



(b) Third approach (2×CO_2 bgc experiment)



(c) Gregory plot for land carbon

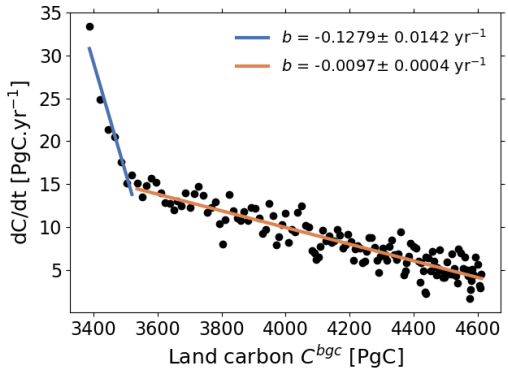


Figure 3.8 : Spectra associated to χ_{NPP} derived with different resolutions and Gregory plot for land carbon. (a) Spectrum derived with the first approach from the 1% bgc experiment; (b) Spectrum derived with the third approach from the 2×CO_2 bgc experiment; (c) "Gregory plot" for land carbon. Values of b indicate the rate at which the time derivative of land carbon changes with respect to the land carbon itself. Ranges of time scales corresponding to each rate accounting for one standard deviation are shown in subfigure (b).

Investigating the two-peak structure and checking its robustness via regional response functions

The reasons for this two-peak structure are conceivable. One possibility is that the different peak time scales originate from the very different characteristic time scales of functionally different elements in the land carbon cycle, such as leaves, wood, litter or soil. Alternatively, it may be that the short time scales reflect the carbon dynamics in the tropics, a region with higher temperatures and thus larger heterotrophic respiration rates (see e.g. Raich and Potter, 1995), while the long time scales may originate from the carbon dynamics in the extra-tropics, where respiration rates are smaller due to lower temperatures. To illustrate how the RFI method may be applied to investigate this question, I analyze in the following the latter hypothesis. For this purpose, I determined separately the linear response functions that characterize the land carbon response to NPP in the tropics and extra-tropics, defined respectively by

$$\Delta C^{bgc,tr}(t) = \int_0^t \chi^{tr}_{NPP}(t-s)\Delta NPP^{tr}(s)ds + \eta^{tr}(t), \tag{3.10}$$

$$\Delta C^{bgc,et}(t) = \int_0^t \chi^{et}_{NPP}(t-s)\Delta NPP^{et}(s)ds + \eta^{et}(t). \tag{3.11}$$

Tropics were defined as the region between latitudes 30° south and 30° north, and extra-tropics as the remaining part of the globe. The data was taken once more from the $2\times CO_2$ bgc experiment. Results are shown in Fig. 3.9a. As seen, the response function for the tropics $\chi^{tr}_{NPP}(t)$ decays faster towards zero than for the extra-tropics $\chi^{et}_{NPP}(t)$. Interpreting the response functions as the impulse response of the land carbon, this means that after an impulse, carbon is indeed respired faster in the tropics than in the extra-tropics. Thus, in principle the response functions are in agreement with our hypothesis.

Before investigating the spectrum, I assess the robustness of the recovered response functions and their spectra by performing two different tests. In the first test, I show that the global response function χ_{NPP} can be reconstructed by combining χ^{tr}_{NPP} and χ^{et}_{NPP}. For this purpose, I write the global land carbon as the sum of the land carbon in the tropics and extra-tropics:

$$\Delta C^{bgc}(t) = \Delta C^{bgc,tr}(t) + \Delta C^{bgc,et}(t). \tag{3.12}$$

Plugging (3.5), (3.10) and (3.11) in (3.12) and recognizing that $\eta(t) = \eta^{tr}(t) + \eta^{et}(t)$ gives

$$\int_0^t \chi_{NPP}(t-s)\Delta NPP(s)ds = \int_0^t \chi^{tr}_{NPP}(t-s)\Delta NPP^{tr}(s)ds + \int_0^t \chi^{et}_{NPP}(t-s)\Delta NPP^{et}(s)ds. \tag{3.13}$$

Applying a Laplace transform to both sides of (3.13) and reorganizing the resulting equation gives

$$\widetilde{\chi}_{NPP}(p) = \frac{\widetilde{\chi^{tr}}_{NPP}(p)\widetilde{\Delta NPP^{tr}}(p) + \widetilde{\chi^{et}}_{NPP}(p)\widetilde{\Delta NPP^{et}}(p)}{\widetilde{\Delta NPP}(p)}. \tag{3.14}$$

Therefore, the Laplace transformed $\widetilde{\chi}_{NPP}(p)$ can be obtained from combining the NPP forcings and the response functions of the tropics and extra-tropics. Hence, if the response

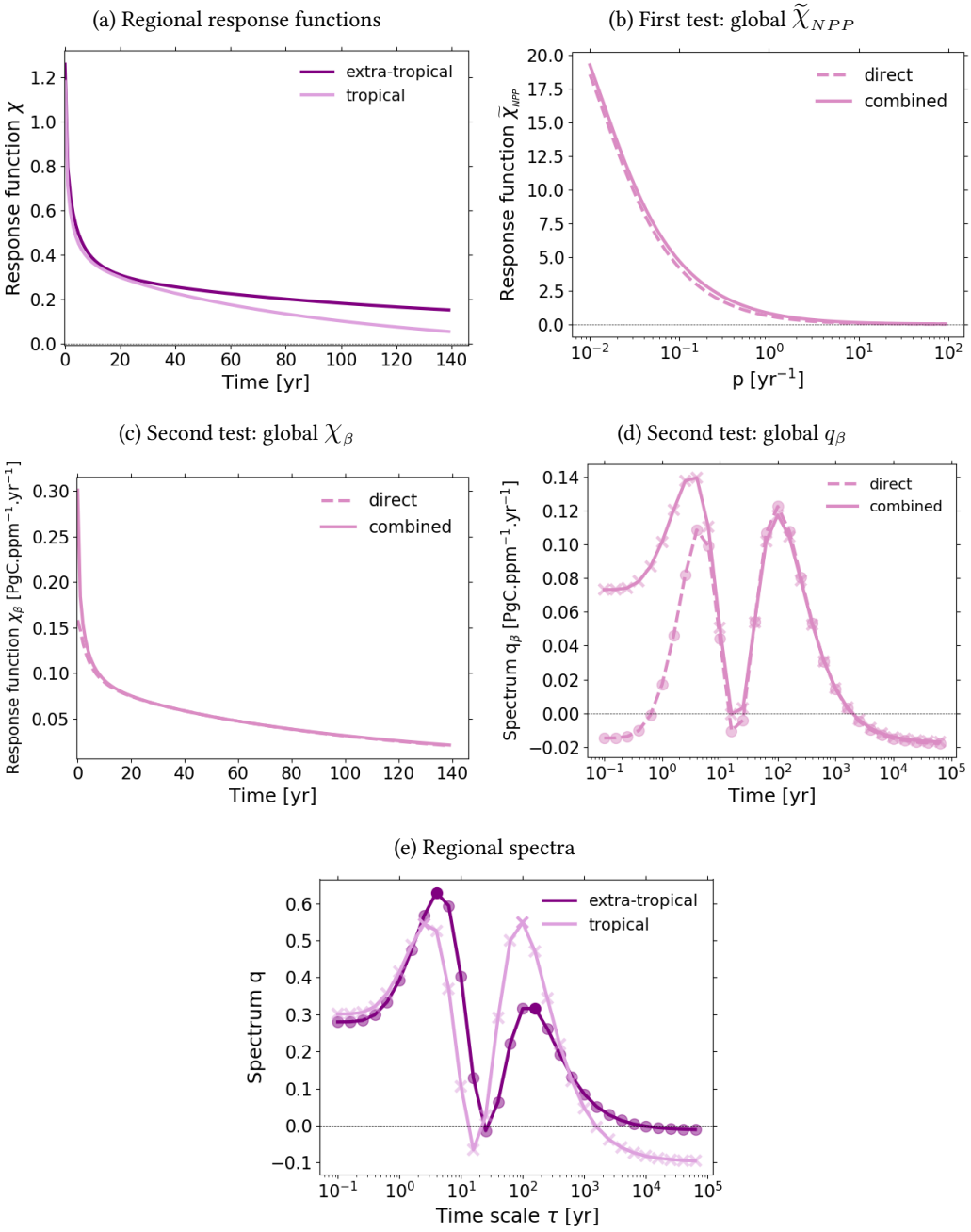


Figure 3.9 : Investigation of the land carbon response in the tropics and extra-tropics and how the regional response functions combine to the global response functions. The analysis is based on the 2×CO_2 bgc experiment. (a) Regional response functions; (b) Laplace transform $\widetilde{\chi}_{NPP}$ of global $\widetilde{\chi}_{NPP}$ obtained directly from the global carbon response (solid line) and from combining the tropical and extra-tropical response functions (dashed line); (c) As (b) but for χ_β; (d) As (c) but for q_β; (e) Tropical and extra-tropical spectra. For more details see text.

functions are correctly recovered by the RFI algorithm, eq. (3.14) should hold at least approximately. In order to check this, I computed the Laplace transforms analytically by approximating the NPP forcings by step functions (since we take the $2\times CO_2$ bgc experiment), and using eq. (2.12) for the response functions. Fig. 3.9b shows the quality of the result. The small discrepancy between $\widetilde{\chi}_{NPP}(p)$ obtained from the global carbon response and from combining the regional responses can at least partially be explained by the approximation error made to represent the forcings by a step function.

As a second test, I show that also the global response function $\chi_\beta(t)$ and its associated spectrum $q_\beta(\tau)$ can be obtained from combining $\chi^{tr}_{NPP}(t)$ and $\chi^{et}_{NPP}(t)$. To this end, I first obtain the tropical $\chi^{tr}_\beta(t)$ and extra-tropical $\chi^{et}_\beta(t)$ by applying (3.9):

$$\chi^{tr}_\beta(t) = \frac{\partial NPP^{tr}}{\partial c}\Big|_{c=c_{PI}} \chi^{tr}_{NPP}(t), \tag{3.15}$$

$$\chi^{et}_\beta(t) = \frac{\partial NPP^{et}}{\partial c}\Big|_{c=c_{PI}} \chi^{et}_{NPP}(t). \tag{3.16}$$

Using the obtained response functions, one can now write (3.3) for global, tropical and extra-tropical carbon. Plugging the result into (3.12) gives

$$\int_0^t \chi_\beta(t-s)\Delta c(s)ds = \int_0^t [\chi^{tr}_\beta(t-s) + \chi^{et}_\beta(t-s)]\Delta c(s)ds. \tag{3.17}$$

Hence, one can infer that

$$\chi_\beta(t) = \chi^{tr}_\beta(t) + \chi^{et}_\beta(t). \tag{3.18}$$

Therefore, the global response function $\chi_\beta(t)$ can be obtained from $\chi^{tr}_\beta(t)$ and $\chi^{et}_\beta(t)$. Also, since $\chi(t)$ is given by eq. (2.12), eq. (3.18) implies that one can also obtain the global spectrum $q_\beta(\tau)$ from combining the regional spectra:

$$q_\beta(\tau) = q^{tr}_\beta(\tau) + q^{et}_\beta(\tau). \tag{3.19}$$

Therefore, using the recovered response functions for tropical and extra-tropical carbon one can obtain the global response function χ_β and its associated spectrum q_β. Accordingly, in this second test I check equations (3.18) and (3.19). For the calculation of the derivatives in (3.15) and (3.16) I fitted $NPP^{tr} = NPP^{tr}(c)$ and $NPP^{et} = NPP^{et}(c)$ once again by a polynomial of order 6 (which obtained the best results for global NPP in Fig. 3.6b) and took the derivatives from the fits. For $\chi_\beta(t)$ I used the recovery with best quality from Fig. 3.6 ("$\text{NPP}^{2\times}_{pol6}$"). The spectra $q_\beta(\tau)$, $q^{tr}_\beta(\tau)$ and $q^{et}_\beta(\tau)$ are obtained by scaling the spectra from $\chi_{NPP}(t)$, $\chi^{tr}_{NPP}(t)$ and $\chi^{et}_{NPP}(t)$ by the respective derivatives $\frac{\partial NPP}{\partial c}\Big|_{c=c_{PI}}$, $\frac{\partial NPP^{tr}}{\partial c}\Big|_{c=c_{PI}}$ and $\frac{\partial NPP^{et}}{\partial c}\Big|_{c=c_{PI}}$. Results are shown for χ_β in subfigure (c) and for q_β in subfigure (d). As seen in Fig. 3.9c, the reconstruction matches almost perfectly the values of χ_β for times beyond about 20 years. Likewise, the spectrum q_β is almost perfectly reconstructed for time scales above 6 years, with a slightly larger discrepancy between 15 and 25 years (see Fig. 3.9d). A larger disagreement is seen below 20 years for χ_β and

below 6 years for q_β. One of the reasons is that only little information is available for time scales smaller than 1 year because this is the time step taken for the data. However, Appendix B.1 shows that time scales much shorter than the time step affect only $\chi(0)$. The main reason for the disagreement is that high frequencies (and thus small time scales) are the most problematic to recover due to the ill-posedness of the problem that obscures the signal particularly at high frequencies (compare Figs. 2.6a$_1$ and 2.6b$_1$). Despite the disagreement at small time scales, subfigures (c) and (d) add confidence that: 1) The response functions for global, tropical and extra-tropical carbon can be trusted over mid-to-long time scales (subfigure (c)); and 2) the two-peak spectrum obtained from global land carbon indeed characterizes the response, since its computation from two independent approaches (combining regional spectra in subfigure (d) and the Gregory plot in Fig. 3.8c) yield similar results with characteristic time scales matching the peaks of the spectrum.

To check whether the hypothesis of the tropical / extra-tropical origin of the two time scales indeed explains the two peaks in the global spectrum, I show in Fig. 3.9e the spectra of χ^{tr}_{NPP} and χ^{et}_{NPP}. Contrary to the hypothesis, the two-peak structure is found in both spectra and even more surprisingly the long time scales (around 100 years) have a stronger weight in the tropics than the extra-tropics. In summary, the short and long time-scale peaks are not explained by different tropical and extra-tropical carbon dynamics.

Nevertheless, it is still interesting that the same structure arises in both spectra, since it provides insight into the temporal structure of the land carbon dynamics in these two regions. The differences in the spectra must explain for instance the discrepancy between the tropical and extra-tropical response functions seen in Fig. 3.9a. However, visual inspection of subfigure (e) does not help determining which are the main time scales that explain the discrepancy. An alternative strategy is to modify one spectrum by replacing some of its values by the corresponding values from the other, and then check the result in the modified response function. Results of such procedure are shown in Fig. 3.10: In subfigure (a) I modify the spectrum for tropical carbon by substituting its values from $\tau = 63$ years on by the corresponding values for the extra-tropical spectrum. In subfigure (b) I show the resulting modified response function: It matches almost entirely the extra-tropical one. This indicates that the discrepancy is associated to contributions to the response acting on time scales above $\tau = 63$ years.

To study this discrepancy in more detail, one may separate it in a difference in the offset of the response functions, i.e. the difference between the last values at which they are computed, and a difference in the rate of decay, which starts to arise approximately from 20 years. By substituting in the tropical spectrum only values from $\tau = 63$ to $\tau = 251$ years (subfigure (c)), we see that the resulting response function (subfigure (d)) looks exactly like the extra-tropical one with a difference in the offset. By adding this difference to the modified tropical response function, it matches once again the extra-tropical one. Hence, we see that the difference in the decay rate of the two response functions is related to time scales between $\tau = 63$ to $\tau = 251$ years, while the difference in the offset is related to time scales longer than $\tau = 251$ years. This demonstrates that the difference in the decay of the response functions is to a large extent explained by time scales within or not much longer

than the range of time scales covered by data (between 1 and 140 years, as explained in the discussion of Fig. 3.8b), which adds confidence that the observed discrepancy is robust. The fact that long time scales outside this range modify mostly the offset is explained in Appendix B.1.

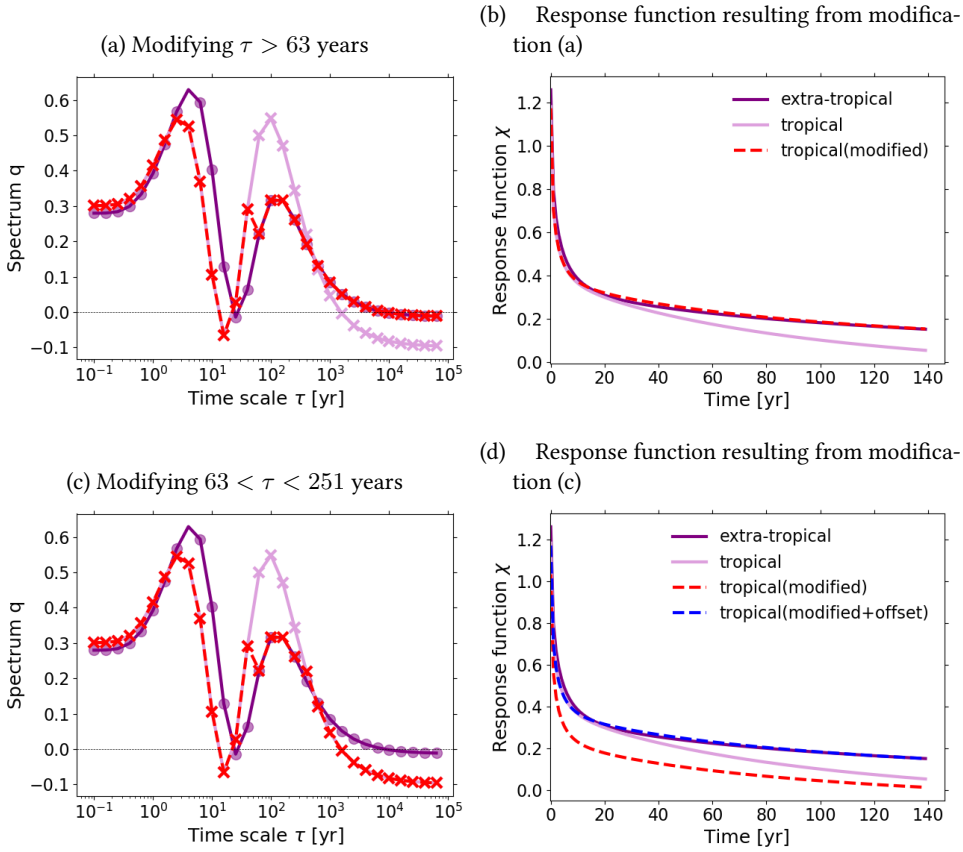


Figure 3.10 : Investigation of differences in the response functions for tropics and extra-tropics. (a) Modified tropical spectrum by taking values of the extra-tropical spectrum from $\tau = 63$ years onwards; (b) resulting modified response function; (c) Modified tropical spectrum by taking values of the extra-tropical spectrum from $\tau = 63$ to $\tau = 251$ years; (d) resulting modified response function. First row shows that the difference in the response functions is associated to time scales greater than 63 years. The second row shows that the difference in the decay is related to time scales between 63 and 251 years, while time scales greater than 251 years change mostly the offset.

3.6 SUMMARY AND DISCUSSION

Although the α-β-γ formalism provides a useful framework for model intercomparison, it only characterizes the sensitivity of a model for a particular perturbation. Instead, one would like to characterize the sensitivity as such. As discussed in section 2.2, the dependence of the α-β-γ values on the considered perturbation scenario arises because of the internal memory of the system, i.e. because of the dependence of the response on past values of the perturbation. When the memory is taken into account, linear response functions arise as natural generalization of the α-β-γ sensitivities. But a fundamental step for applying this generalized framework is to identify the appropriate linear response functions. In chapter 2, I developed a method to identify linear response functions from data using only information from an arbitrary perturbation experiment and a control simulation. In that chapter, the robustness of the method in the presence of noise and nonlinearity was demonstrated in applications to data from perturbation experiments performed with a toy model.

Generalized land carbon sensitivities in the MPI-ESM

In the present chapter, I demonstrated that the developed method (RFI method) can as well be employed to derive response functions from C^4MIP data. Here, I identified for the MPI-ESM, using data from standard 1% experiments, the land carbon generalized sensitivities χ_γ and χ_β, i.e. the linear response functions that generalize the γ- and β-sensitivities for land carbon. The robustness of the identified generalized sensitivities was demonstrated by their ability to predict the response from experiments not used for the identification. With the aid of additional experiments, I estimated the linear regime that gives the range of forcing strengths for which each generalized sensitivity can predict model responses. For χ_γ, results indicate a linear response at least until the end of the 1% experiment, corresponding to temperature perturbations of around 6 K. For χ_β, the estimate is for CO_2 perturbation strengths up to 100 ppm. In addition, I analyzed different approaches that may be employed to improve the quality of the recovery of the response functions. For χ_γ, taking the response from a $2\times CO_2$ experiment demonstrated to give smaller prediction errors for every response evaluated, suggesting that this type of experiment gives also a better recovery. For χ_β, I used the knowledge of the relationship between NPP and CO_2 to account for some of the nonlinearities in the response. I found that nonlinearities can to a large extent be explained by the nonlinear relationship between NPP and CO_2: By using NPP instead of CO_2 as forcing an approximately linear response is found over the whole 1% experiment, i.e. until CO_2 perturbations of about 850 ppm. In addition, this approach yielded the best recovery for χ_β. Evidences for this conclusion are the quality of its prediction and the detailed spectrum it yields for the response.

Spectrum of land carbon time scales

Obtaining the spectrum of the response is an additional advantage of the RFI method. Most methods in the literature either recover $\chi(t)$ pointwise – and therefore do not give the spectrum –, or try to fit it with few exponents without accounting for ill-posedness, which does not give reliable results for the spectrum (see the famous example by Lanczos, 1956). In the application to the MPI-ESM, obtaining the spectrum has proven advantageous for two reasons. First, it allows to formally extend the recovery of the response function to any time range: While it is unrealistic to expect to recover from a time series information on all time scales, results from such an extension of the response function (Figs. 3.7a and b) demonstrated that the recovered response function nevertheless contains information on times reasonably longer than the time series length used for the recovery.

Second, it gives valuable insight into the internal dynamics of the system. In particular, for the application here the spectrum gives the most relevant time scales for the land carbon response on a global or regional level. The spectrum associated with the best recovery of χ_β showed two peak time scales for the global response: One around 4 and another around 100 years. I illustrated how the reasons for such global spectrum can be investigated by analyzing the hypothesis whether the different peak time scales could be related to different land carbon responses over tropics and extra-tropics. To first obtain evidence for the robustness of the results, I showed that it is possible to reconstruct the global spectrum from the recovered tropical and extra-tropical spectra (see Fig. 3.9d), and that similar time scale ranges can be obtained via a "Gregory plot" approach (see Fig. 3.8b and c). Further, I demonstrated that the recovered tropical and extra-tropical response functions combine to the identified global response functions indicating consistency between regional and global recovery (see Fig. 3.9b and c). Differences between the tropical and extra-tropical response functions showed that indeed land carbon responds at faster time scales in the tropics than in the extra-tropics. Analyses of the obtained spectra revealed that the difference in the decay of the regional response functions lies on time scales between around 60 and 250 years, with time scales above around 250 years contributing mostly to the offset between the response functions (Fig. 3.10). Nevertheless, the same type of two-peak spectrum was found for the two regional responses, suggesting that the hypothesis on their separate regional origin may be rejected.

Results for the spectrum should be always investigated with care, since deriving the spectrum is a more complicated problem than deriving the response function. This is partially explained by the degree of ill-posedness of each problem: Deriving the response function one is solving a deconvolution problem (e.g. Hansen, 2002), while by deriving the spectrum one is solving in addition a type of inverse Laplace transform problem (to obtain the spectrum from the response function (2.12)), known to be extremely ill-posed (Istratov and Vyvenko, 1999). This difficulty was exemplified by results from appendices B.1 and B.2, discussing cases where the response function can be perfectly recovered but the recovered spectrum is relatively poor.

As mentioned in subsection 2.3.1, one of the advantages of assuming the ansatz (2.11) instead of the widely used (2.10) is that (2.11) better describes our expectation that variables integrated over many different climate zones present a large spectrum of time scales. Nevertheless, as shown in Appendix B.2, even if the underlying spectrum turns out to be a discrete spectrum with only few time scales, for a sufficiently good signal-to-noise ratio the ansatz (2.11) may still give a reasonable approximation.

Final remarks

In this study I investigated the generalization of the climate-carbon cycle sensitivities, which solves the scenario dependence problem noted e.g. by Gregory et al., 2009 and Arora et al., 2013 to linear order in the perturbation – an approach that in principle can be extended to higher orders (Ruelle, 1998b, Lucarini, 2009). The main challenge in following this idea is to identify from published C^4MIP data the generalized climate-carbon cycle sensitivities. Applying the RFI method I obtained for the MPI-ESM using C^4MIP 1% simulations the generalized land carbon sensitivities. The robustness of the obtained sensitivities suggests that the RFI method is appropriate to identify from published C^4MIP data also the generalized climate-carbon cycle sensitivities for other models. In addition, the estimates for the linear regime obtained by employing the recovered sensitivities to predict additional experiments give insight into the range of perturbation strengths for which this generalized framework might be valid for other models, for which data from additional experiments are not available. Assuming that the linear regime estimates for the MPI-ESM extend to other models, the generalized sensitivity χ_γ characterizes the sensitivity of the land carbon to temperature perturbations until at least 6 K. As a consequence, in models with temperature response similar or weaker than the MPI-ESM, χ_γ can be recovered taking the full time series from the 1% experiment. On the other hand, the generalized sensitivity χ_β characterizes the sensitivity of land carbon to CO_2 perturbations until around 100 ppm. Since in 1% experiments the CO_2 perturbation is prescribed, χ_β can be recovered taking data for the first 30 years. Alternatively, as discussed in subsection 3.4, the time range for the recovery of χ_β can be extended if one takes as forcing NPP instead of CO_2 (third approach investigated in subsection 3.4). For models with NPP response similar or weaker than the MPI-ESM, by this approach χ_β can be recovered taking the full time series from the 1% experiment. Still, as the relatively successful predictions in Fig. 3.7b suggest, even if χ_β is recovered by taking CO_2 as forcing for only 30 years of the 1% experiment (first approach investigated in subsection 3.4), it may still have sufficient information to predict responses of the model for a time range of 140 years.

All that said, it is worth noting that the recovered generalized sensitivities are only an approximation (see subsections 3.3 and 3.4). The quality of the recovery depends on the quality of the data, i.e. the SNR and the range of frequencies excited by the perturbation (see the Laplace transform analysis in subsection 3.3). Therefore, despite the several tests where the recovered response functions demonstrated to correctly predict different experiments, it is not clear whether the quality of these response functions is sufficiently good to predict

really any type of experiment. At first sight it may seem that this contradicts the claim that response functions generalize Friedlingstein's sensitivities by giving the response to any type of perturbation (see section 2.2). But this issue is different from the scenario dependence encountered in the α-β-γ framework: It is not a fundamental issue underlying the formulation of the generalized sensitivities, but only a technical issue related to the quality of the data used for the recovery. Therefore, in principle the quality of the recovered generalized sensitivities – and therefore their ability to predict experiments – can be improved by enhancing the quality of the data (for instance by taking ensemble averages as in Ragone et al., 2016 or perturbation experiments that cover wider ranges of frequencies).

4

CLIMATE-CARBON CYCLE FEEDBACKS FOR WEAK PERTURBATIONS IN CMIP5 MODELS

4.1 INTRODUCTION

Friedlingstein's α-β-γ framework (Friedlingstein et al., 2003) was developed to separately quantify the so-called climate-carbon cycle feedbacks. Such feedbacks are processes that arise within the land and ocean carbon cycle as a response to perturbations in atmospheric CO_2, and act either taking up or releasing CO_2 from/to the atmosphere in addition to the initial perturbation. For example, an increase in atmospheric CO_2 leads to warming; higher temperatures result in larger heterotrophic respiration rates (Raich and Potter, 1995); in turn, such larger rates lead to an additional release of CO_2 to the atmosphere, thereby amplifying the initial CO_2 increase. The interplay between these feedbacks determine the net climate-carbon cycle feedback, which affects the CO_2 airborne fraction and, thereby, the climate. Since its formulation, Friedlingstein's framework has been widely applied in the study of the global carbon cycle and its interaction with climate (Friedlingstein et al., 2006, Gregory et al., 2009, Arneth et al., 2010, Zickfeld et al., 2011, Boer and Arora, 2013, Arora et al., 2013, Schwinger et al., 2014, Friedlingstein et al., 2014, Adloff et al., 2018, Williams et al., 2019, Goodwin et al., 2019, Jones and Friedlingstein, 2020). Nevertheless, it is well-known that this framework suffers from certain deficiencies. Because it assumes an instantaneous response of the carbon cycle to perturbations (Gregory et al., 2009; see also chapter 2), the resulting quantification of the feedbacks depends on the perturbation scenario considered (Gregory et al., 2009, Boer and Arora, 2013, Arora et al., 2013). Additionally, because it is a static framework where usually the feedback strengths are calculated by selecting a particular time period of a particular perturbation experiment, it does not give insight into the dynamics of the feedback mechanisms.

Recently, studies have proposed a generalization of Friedlingstein's framework that accounts for the memory of the carbon cycle and thereby gives a consistent quantification of the feedbacks (Rubino et al., 2016, Enting and Clisby, 2019). The approach is to consider weak perturbations and quantify the feedbacks by means of linear response functions. These functions give rise to a description of the linear dynamics of the global carbon cycle in terms of the feedbacks at different time scales. From this description one can as well determine a time-scale dependent form of the airborne fraction. In this generalization, both the climate-carbon cycle feedbacks and the airborne fraction are uniquely characterized for all perturbation scenarios. Although this generalization has been proposed, no attempt has been made so far to systematically investigate its applicability to Earth System Models.

The goal of the present study is twofold. First, to demonstrate that the description of the carbon cycle dynamics given by this generalization is indeed consistent with the behaviour of the carbon cycle in an Earth System Model. Second, to apply this framework to quantify in different Earth System Models the climate-carbon cycle feedbacks and the airborne fraction. Since the land carbon sensitivities were already derived for the MPI-ESM in the last chapter, the test of the consistency of the framework will be performed as well with the MPI-ESM. As will be seen, the test gives a positive result, which encourages the application of the framework to quantify the feedbacks and the airborne fraction in other Earth System Models. For this application I select other models participating in the fifth phase of the Coupled Model Intercomparison Project (Taylor et al., 2012), or CMIP5 models (see a description in Arora et al., 2013, Dunne et al., 2013). The analysis in the present study will show that one can understand the dynamics of the airborne fraction from the behaviour of the climate-carbon cycle feedbacks, and that it is possible to pinpoint the feedback that mostly contributes to the observed model spread in the airborne fraction at different time scales, which is the land biogeochemical feedback (or land carbon-concentration feedback).

The outline of the chapter is as follows. In next section, the generalized α-β-γ framework is introduced. For the derivation of the feedbacks and the airborne fraction, one needs to derive the generalized sensitivities. Since the land carbon generalized sensitivities in the MPI-ESM have been derived in the last chapter, in section 4.3 the ocean carbon and temperature sensitivities are derived. In section 4.3 it is also shown that these sensitivities predict the separate responses of the carbon cycle to different CO_2 perturbation scenarios (concentration-driven simulations). In section 4.4, it is demonstrated that the generalized framework correctly predicts also the response of the coupled carbon cycle in the MPI-ESM as observed from emission-driven simulations. Together, results from chapter 3, section 4.3 and section 4.4 demonstrate that the generalized framework indeed gives a consistent description of the linear dynamics of the global carbon cycle in the MPI-ESM. Finally, in section 4.5 the feedbacks and airborne fraction are derived and analyzed for the different CMIP5 models. In sections 4.6 and 4.7 the results are summarized and discussed. Additional calculations are shifted to the appendices.

4.2 GENERALIZED α-β-γ FRAMEWORK

The purpose of this study is to demonstrate that the generalized α-β-γ framework (Rubino et al., 2016, Enting and Clisby, 2019) consistently describes the linear dynamics of the global carbon cycle in an Earth System Model – the MPI-ESM –, and to employ this framework to quantify the climate-carbon cycle feedbacks and airborne fraction in different CMIP5 models. To prepare for the subsequent investigation I introduce in this section the generalized α-β-γ framework.

The goal of the α-β-γ framework originally introduced by Friedlingstein et al., 2003 is to disentangle the strength of climate-carbon cycle feedbacks into contributions arising from changes in atmospheric CO_2 and climate. The basic idea underlying the framework is that

the carbon cycle response to CO_2 perturbations can be split into a biogeochemical response, which results from purely changing atmospheric CO_2 concentrations, and a radiative response, which results from the climate change that arises from the CO_2 perturbations. In the original framework these two types of responses are characterized by the β- and γ-sensitivities. In addition, climate change is represented by temperature, whose sensitivity to CO_2 perturbations is characterized by α.

Although this framework gives insight into the main effects of atmospheric CO_2 perturbations onto the global carbon cycle, the α-β-γ sensitivities characterize the carbon cycle response only for a particular perturbation scenario (Gregory et al., 2009, Arora et al., 2013). As discussed in section 2.2, for sufficiently weak perturbations these sensitivities can be generalized by considering linear response functions, which instead characterize the response to any type of perturbation scenario. The framework studied in this chapter arises from this generalization. To introduce it, I start from the carbon balance equation that couples the different subsystems of the global carbon cycle:

$$\Delta C_A(t) = \int_0^t E(s)ds - \Delta C_L(t) - \Delta C_O(t), \tag{4.1}$$

where ΔC_A, ΔC_L and ΔC_O are changes in global carbon content in the atmosphere, land and ocean with respect to the pre-industrial equilibrium state, and $E(t)$ is the time-dependent flux of carbon emissions. Following the framework from Friedlingstein et al., 2003, the land and ocean carbon changes in (4.1) are assumed to depend on the CO_2 concentration (characterizing the biogeochemical response) and on temperature (characterizing the radiative response). For weak perturbations, in the framework of linear response (see section 2.2) these changes can be understood as responses, while changes in CO_2 and temperature can be understood as perturbations, giving the following equations[1]:

$$\Delta C_L(t) = \int_0^t \chi_\beta^{(L)}(t-s)\Delta c(s)ds + \int_0^t \chi_\gamma^{(L)}(t-s)\Delta T_L(s)ds, \tag{4.2}$$

$$\Delta C_O(t) = \int_0^t \chi_\beta^{(O)}(t-s)\Delta c(s)ds + \int_0^t \chi_\gamma^{(O)}(t-s)\Delta T_O(s)ds, \tag{4.3}$$

where Δc is the change in CO_2 concentration with respect to the pre-industrial state, ΔT_L and ΔT_O are changes in land and ocean global mean near-surface air temperature with respect to the pre-industrial state, and $\chi_\beta^{(L)}$, $\chi_\gamma^{(L)}$, $\chi_\beta^{(O)}$, and $\chi_\gamma^{(O)}$ are the linear response functions that generalize the sensitivities $\beta^{(L)}$, $\gamma^{(L)}$, $\beta^{(O)}$, and $\gamma^{(O)}$. Additionally, the temperature can be understood as a response to weak CO_2 perturbations:

$$\Delta T_L(t) = \int_0^t \chi_\alpha^{(L)}(t-s)\Delta c(s)ds, \tag{4.4}$$

$$\Delta T_O(t) = \int_0^t \chi_\alpha^{(O)}(t-s)\Delta c(s)ds, \tag{4.5}$$

1 The framework of linear response describes only the response in the *ensemble average* of a particular observable. Therefore strictly one had to write the response formulas adding a noise term $\eta(t)$ as in eq. (2.9). But to keep calculations simple it is assumed that the noise term is always sufficiently small so that it can be ignored.

where $\chi_\alpha^{(L)}$ and $\chi_\alpha^{(O)}$ are the linear response functions that generalize the sensitivities $\alpha^{(L)}$ and $\alpha^{(O)}$. The response functions that generalize the α-β-γ sensitivities are referred to as *generalized sensitivities*.

Note that in principle, assuming that the land and ocean carbon cycle indeed have a linear response and that the response can be described in terms of the dependencies on CO_2 and temperature, by inserting eqs. (4.4) and (4.5) into (4.2) and (4.3) one can describe the separate land and ocean carbon responses to any temporal development of weak atmospheric CO_2 perturbations. Nevertheless, in reality the land and ocean carbon cycle form a coupled system that responds to CO_2 emissions. Therefore, to study the response of this coupled system one must insert eqs. (4.2), (4.3), (4.4), and (4.5) into (4.1). Because convolution in the time domain is equivalent to multiplication in the Laplace domain, calculations are facilitated by Laplace-transforming the resulting equation. After these two operations eq. (4.1) becomes

$$\Delta\widetilde{c}(p) = \frac{\widetilde{e}(p)}{p} - \left[\widetilde{\chi}_\beta^{(L)}(p) + \widetilde{\chi}_\gamma^{(L)}(p)\widetilde{\chi}_\alpha^{(L)}(p) + \widetilde{\chi}_\beta^{(O)}(p) + \widetilde{\chi}_\gamma^{(O)}(p)\widetilde{\chi}_\alpha^{(O)}(p)\right]\frac{\Delta\widetilde{c}(p)}{k}, \quad (4.6)$$

where the tilde denotes Laplace-transformed quantities and $\Delta\widetilde{C}_A(p) = k\,\Delta\widetilde{c}(p)$, $\widetilde{E}(p) =: k\,\widetilde{e}(p)$ with $k = 2.12\ PgC/ppm(CO_2)$ (Ciais et al., 2013, p. 417) were used to convert units in terms of PgC to units in terms of $ppm(CO_2)$. In this way, the carbon balance equation is not anymore described in the time domain but in terms of the rates p, whose inverses $1/p$ can be understood as time scales. From eq. (4.6) it is clear that the atmospheric CO_2 response $\Delta\widetilde{c}(p)$ depends not only on the cumulated emissions $\widetilde{e}(p)/p$ but also on a fraction of itself determined by the sum in the brackets divided by k. This additional input to the response that depends on the response itself characterizes the net climate-carbon cycle feedback (Roe, 2009). Following Roe, 2009, Gregory et al., 2009 and Adloff et al., 2018, eq. (4.6) can be rewritten as

$$\Delta\widetilde{c}(p) = \widetilde{A}(p)\frac{\widetilde{e}(p)}{p}, \quad (4.7)$$

where

$$\widetilde{A}(p) := \frac{1}{1-\widetilde{f}(p)}, \quad (4.8)$$

with

$$\widetilde{f}(p) := \widetilde{f}_\beta^{(L)}(p) + \widetilde{f}_{\gamma\alpha}^{(L)}(p) + \widetilde{f}_\beta^{(O)}(p) + \widetilde{f}_{\gamma\alpha}^{(O)}(p), \quad (4.9)$$

and

$$\widetilde{f}_\beta^{(L)}(p) := -\frac{\widetilde{\chi}_\beta^{(L)}(p)}{k}, \quad \widetilde{f}_{\gamma\alpha}^{(L)}(p) := -\frac{\widetilde{\chi}_\gamma^{(L)}(p)\widetilde{\chi}_\alpha^{(L)}(p)}{k}, \quad (4.10)$$

$$\widetilde{f}_\beta^{(O)}(p) := -\frac{\widetilde{\chi}_\beta^{(O)}(p)}{k}, \quad \widetilde{f}_{\gamma\alpha}^{(O)}(p) := -\frac{\widetilde{\chi}_\gamma^{(O)}(p)\widetilde{\chi}_\alpha^{(O)}(p)}{k}. \quad (4.11)$$

The function $\widetilde{A}(p)$ is the airborne fraction, which in this generalized framework determines the atmospheric CO_2 response to any type of emissions scenario $\widetilde{e}(p)$ including the

effect of climate-carbon cycle feedbacks. As explained in Adloff et al., 2018 and Roe, 2009, from the point of view of feedback analysis $\widetilde{A}(p)$ is the *gain* of the system: Depending on whether $\widetilde{A}(p)$ is larger or smaller than 1, the inclusion of the feedbacks increases or decreases the response of atmospheric CO_2 at a particular time scale in comparison to a reference system where the atmospheric CO_2 would simply follow the emissions[2]. In other words, depending on whether $\widetilde{A}(p)$ is larger or smaller than 1, the inclusion of the feedbacks results in CO_2 fluxes into or out of the atmosphere in addition to the emissions at a particular time scale.

As shown by eq. (4.8), the airborne fraction $\widetilde{A}(p)$ is completely determined by $\widetilde{f}(p)$. Following the terminology of Roe, 2009 and Adloff et al., 2018, I call $\widetilde{f}(p)$ the *net feedback function*. Depending on whether $\widetilde{f}(p)$ is positive or negative, $\widetilde{A}(p)$ is smaller or larger than 1. Therefore, from the explanation above, a positive or negative $\widetilde{f}(p)$ implies additional CO_2 fluxes into or out of the atmosphere at a particular time scale.

The net feedback function $\widetilde{f}(p)$ is defined in eq. (4.9) as the sum of the feedback functions $\widetilde{f}_{\beta}^{(L)}(p)$, $\widetilde{f}_{\gamma\alpha}^{(L)}(p)$, $\widetilde{f}_{\beta}^{(O)}(p)$ and $\widetilde{f}_{\gamma\alpha}^{(O)}(p)$. These functions quantify at a particular time scale for land and ocean the biogeochemical $\left(\widetilde{f}_{\beta}^{(L)} \text{ and } \widetilde{f}_{\beta}^{(O)}\right)$ and the radiative $\left(\widetilde{f}_{\gamma\alpha}^{(L)} \text{ and } \widetilde{f}_{\gamma\alpha}^{(O)}\right)$ feedback, also known as carbon-concentration and carbon-climate feedback (Arora et al., 2013). The feedback functions, in turn, are determined from the generalized sensitivities via eqs. (4.10) and (4.11). Therefore, the feedbacks, the airborne fraction, and as a consequence the whole linear dynamics of the global carbon cycle depend on the interplay between the different generalized sensitivities. Because eqs. (4.2)-(4.5) determine the response of the land and ocean carbon to atmospheric CO_2 perturbations, by plugging (4.7)-(4.11) into the Laplace-transformed version of (4.2)-(4.5) one can determine also the response of the land and ocean carbon cycle to CO_2 emissions. Hence, in principle eqs. (4.2)-(4.5) and (4.7)-(4.11) completely describe in the time-scale domain the dynamics of the coupled global carbon cycle for any temporal development of weak CO_2 perturbations.

As seen, the generalized sensitivities are the essential ingredient of the generalized α-β-γ framework. Therefore, to achieve the first aim of this study, which is to demonstrate that the generalized framework gives a consistent description of the linear dynamics of the global carbon cycle in the MPI-ESM, one must first derive these sensitivities. The generalized land carbon sensitivities $\chi_{\beta}^{(L)}$ and $\chi_{\gamma}^{(L)}$ were already derived in chapter 3. In the next section I derive the remaining sensitivities. Results from next section together with those from chapter 3 and section 4.4 will demonstrate that the global carbon cycle dynamics described by the generalized framework is consistent with that observed from simulations with the MPI-ESM. This demonstration will add confidence that the framework is appropriate for the application to the CMIP5 models in chapter 4.5.

2 Note that the reference system that I take here is different from that in Friedlingstein et al., 2003. While here I take as reference a system where the atmospheric CO_2 is simply determined by CO_2 emissions without any response from the land and ocean carbon cycle, in Friedlingstein et al., 2003 the reference system includes the biogeochemical response. Therefore eq. (4.9) from this study differs from their eq. (9) not only because I investigate a generalization of their framework, but also because the generalized sensitivities show up in a different manner.

4.3 GENERALIZED SENSITIVITIES FOR OCEAN CARBON AND TEMPERATURE IN THE MPI-ESM

This section complements the results from chapter 3 to derive all generalized sensitivities for the MPI-ESM. Here I derive the generalized sensitivities for the ocean carbon and temperature. I demonstrate that the derived generalized sensitivities characterize the separate biogeochemical and radiative responses of ocean carbon and also the response of land and ocean temperature to different temporal developments of atmospheric CO_2. Insofar this section prepares for applying the generalized framework to describe the dynamics of the global carbon cycle as a whole in section 4.4. In addition, results from this section together with those from chapter 3 provide the approaches that will be used in sections 4.4 and 4.5 to derive the generalized sensitivities for all CMIP5 models. These results also give the linear regime estimates that determine for the MPI-ESM the range of perturbation strengths for which the generalized framework is applicable, thereby shedding some light into what this range might be also for other models.

To derive the generalized sensitivities from C^{4}MIP experiments I employ the RFI method (chapter 2). This method allows for recovering a linear response function $\chi(t)$ taking data from an arbitrary perturbation experiment and a control simulation. The method employs regularization to solve

$$\Delta Y(t) = \int_0^t \chi(t-s)\Delta f(s)ds \tag{4.12}$$

for the response function $\chi(t)$ given data from the response $\Delta Y(t)$ and the perturbation $\Delta f(t)$. In the present section as well as in section 4.5 I take data from standard C^{4}MIP 1% experiments, where atmospheric CO_2 is prescribed to increase at a rate of 1% per year starting from a control run at pre-industrial conditions. Since $\chi_\beta^{(O)}$ characterizes the biogeochemical response of the ocean carbon, I derive it from the 1% experiment performed in the "bgc" (biogeochemically coupled) setup, where only the biogeochemistry of the model is affected by changes in atmospheric CO_2. In contrast, to derive $\chi_\gamma^{(O)}$, $\chi_\alpha^{(O)}$ and $\chi_\alpha^{(L)}$ I use data from the 1% experiment performed in the "rad" (radiatively coupled) setup, where only the radiation code is affected by changes in atmospheric CO_2.

As discussed in chapter 2, the recovery of the response function may be hindered by a poor quality of the data and the presence of strong nonlinearities in the response. Therefore, similarly to chapter 3, in this section I derive the generalized sensitivities by different approaches where I try to account for some of the nonlinearities in the response. The advantage of such approaches is that by accounting for nonlinearities one may take data until higher perturbation strengths and therefore higher signal-to-noise ratios, resulting in a recovery with better quality. The approach that gives the recovery with best quality will be employed to derive the sensitivities for all CMIP5 models in sections 4.4 and 4.5. To estimate the linear regime for which the generalized sensitivities characterize the response, I employ the technique developed in section 3.2. To obtain evidence that the recovered sensitivities indeed characterize the response of the MPI-ESM to any CO_2 perturbation

scenario, I test their ability to predict the response of the model from additional experiments whose data were not used for the recovery (for a description of the experiments see Table 2.1).

4.3.1 *Generalized sensitivity* $\chi_\beta^{(O)}$

Approaches to recover $\chi_\beta^{(O)}$

Similarly to section 3.4, I recover $\chi_\beta^{(O)}(t)$ from two different approaches. In the first approach I consider $\Delta c(t)$ as perturbation and derive $\chi_\beta^{(O)}(t)$ from

$$\Delta C_O^{bgc}(t) = \int_0^t \chi_\beta^{(O)}(t-s)\Delta c(s)ds, \tag{4.13}$$

where $\Delta C_O^{bgc}(t)$ is the ocean carbon response in the "bgc" setup. Because (4.13) is the equation employed in the generalized framework (compare eq. (4.3)), from this first approach I will also obtain an estimate of the linear regime for the ocean biogeochemical response as described in the framework.

In the second approach, I consider the logarithm of c as perturbation and derive $\chi_\beta^{(O)}(t)$ from

$$\Delta C_O^{bgc}(t) = \int_0^t \chi_{\beta,\ln}^{(O)}(t-s)c_{PI}\ln\frac{c(s)}{c_{PI}}ds, \tag{4.14}$$

where c_{PI} is the pre-industrial value for atmospheric CO_2. This second approach is inspired by the fact that nonlinearities in the ocean carbon uptake come mainly from the dissolution of CO_2 in ocean surface water (see e.g. Joos and Bruno, 1996). Because the solubility of additional CO_2 in ocean surface water decreases with rising CO_2 concentration (Hooß et al., 2001), I use a logarithmic representation for the perturbation to try to explicitly describe the nonlinearity between CO_2 concentration and the carbon flux into the ocean. As explained in section 3.4, employing eq. (4.14) has the advantage that $\chi_{\beta,\ln}^{(O)}(t) = \chi_\beta^{(O)}(t)$, i.e. by deriving $\chi_{\beta,\ln}^{(O)}(t)$ from (4.14) one obtains also the desired $\chi_\beta^{(O)}(t)$.

Because changes in atmospheric CO_2 result in a change in the input of carbon into the ocean, Appendix A.4 applies also to $\chi_\beta^{(O)}(t)$, which is therefore monotonic. Hence, for both approaches $\chi_\beta^{(O)}(t)$ is derived enforcing monotonicity by the RFI algorithm (see Fig. 2.2).

Identifying the linear regime of the response

To correctly recover $\chi_\beta^{(O)}(t)$ – and the other response functions in the next subsections – one must first identify the linear regime of the response for which nonlinearities do not

spoil the recovery. To this end I follow the technique presented in section 3.2. The idea behind the technique can be understood as follows. I recover the response function taking data from increasingly longer time periods of the 1% experiment. For each time period that the response function is recovered, I employ it to predict the response of additional 0.5% and 0.75% experiments covering that same time period (for a description of the experiments see Table 2.1). Then, I measure the quality of the recovery of the response function indirectly by the quality of the prediction of the responses. The quality of the prediction is quantified by the prediction error

$$\varepsilon_k := \frac{||\boldsymbol{\Delta}\mathbf{Y}^k - \boldsymbol{\chi} \star \boldsymbol{\Delta}\mathbf{f}^k||}{||\boldsymbol{\Delta}\mathbf{Y}^k||}, \tag{4.15}$$

where $\star$ stands for the convolution operation, $\boldsymbol{\Delta}\mathbf{Y}^k$ and $\boldsymbol{\Delta}\mathbf{f}^k$ are the response and the perturbation in experiment "k" and $\boldsymbol{\chi}$ is the response function recovered from the 1% experiment. Because in the 1% experiment the perturbation strength increases with time, also the signal-to-noise ratio of the data increases. On the other hand, higher perturbation strengths increase nonlinearities. This results in the following trade-off: While a higher signal-to-noise ratio results in a recovery with better quality, larger nonlinearities deteriorate the quality of the recovery. By analyzing the prediction error (4.15) for different perturbation strengths of the 1% experiment one can from this trade-off estimate the linear regime. As exemplified in section 3.2, if the response is approximately linear throughout the range of perturbation strengths of the 1% experiment, the quality of the recovery simply increases with increasing perturbation strengths because of the improvement in the signal-to-noise ratio. Therefore in this case the prediction error decreases with increasing perturbation strengths and the whole range of perturbation strengths of the 1% experiment can be considered within the linear regime. On the other hand, if strong nonlinearities start to arise at some value of the perturbation strength, from this value on the quality of the recovery starts to deteriorate. As a result, from this value on the prediction error starts to increase. In this case, the range of perturbation strengths below this value can be considered an estimate for the linear regime of the response. For a more detailed description of this technique please refer to section 3.2.

Recovery of $\chi_\beta^{(O)}$

As a first step to recover $\chi_\beta^{(O)}(t)$ I therefore evaluate the linear regime of the biogeochemical response of ocean carbon by the technique described above. To this end, I derived $\chi_\beta^{(O)}(t)$ from eq. (4.13) – first approach described at the beginning of this subsection – taking data from the 1% bgc experiment and employed this same equation to predict the response from the 0.5% and 0.75% bgc experiments (see Table 2.1). Fig. 4.1a shows the resulting prediction error (see eq. (4.15)) as a function of the final forcing strength in the 1% bgc experiment. Similarly as for the biogeochemical response of land carbon (see chapter 3), minima are found for final forcing strengths below 100 ppm (about 94 ppm for the 0.75% and 58 ppm for the 0.5% bgc experiment), indicating the presence of strong nonlinearities for forcing

strengths above this value. Because both minima are "flat", i.e. the error does not change substantially for values around the minima, I take as an estimate for the linear regime forcing strengths below the highest minimum – about 94 ppm in the 0.75% curve.

To try to improve the quality of the recovery, $\chi_\beta^{(O)}(t)$ was derived also by the second approach (eq. (4.14)). To understand if the quality of the recovery can indeed be improved by this approach, one must check if eq. (4.14) indeed accounts for some of the nonlinearities in the response. If this is the case, eq. (4.14) should predict the 0.5% and 0.75% bgc responses better than eq. (4.13). To check this, I recovered $\chi_\beta^{(O)}(t)$ by this second approach and employed the recovered $\chi_\beta^{(O)}(t)$ in eq. (4.14) to predict these responses. The resulting prediction error is shown Fig. 4.1b against the final forcing strength in the 1% bgc experiment. As seen, overall the error is substantially smaller than in subfigure (a). Minima are still found, but now at values between 100 and 200 ppm and with smaller optimal errors. In contrast to subfigure (a), after the minima the error increases only slightly for increasing final forcing strength. Therefore, eq. (4.14) seems to indeed account for some of the nonlinearities in the response. This confirms the expectation that nonlinearities in the ocean carbon response to CO_2 arise mainly from the dissolution of CO_2 in the ocean (see explanation after eq. (4.14)). Hence, with this approach one can derive $\chi_\beta^{(O)}(t)$ taking data from the 1% bgc experiment until larger perturbation strengths and therefore higher signal-to-noise ratios than with the first approach, making it in principle possible to recover $\chi_\beta^{(O)}(t)$ with better quality.

But despite the overall reduction in the prediction error, Fig. 4.1b still shows a slightly increasing trend in the prediction error after the minima, indicating the presence of possibly non-negligible nonlinearities in the response. Therefore I check whether the second approach indeed gives a better recovery for $\chi_\beta^{(O)}(t)$ by comparing the recovery from the second approach with that from the first approach. Fig. 4.1c shows the response function recovered with each approach taking data from the 1% bgc experiment until the optimal final forcing strength (I take 94 ppm or 30 years of the 1% bgc experiment for the first approach and 138 ppm or 50 years of the 1% bgc experiment for the second) and the response function recovered with the second approach but taking data until the maximal final forcing strength in the 1% bgc experiment, i.e. the whole time series (since for those forcing strengths the error in Fig. 4.1b is still small). As seen, all recovered response functions are surprisingly similar. This similarity is most likely due to the combination of an overall high signal-to-noise ratio of the ocean biogeochemical response (as indicated by the small errors even at small perturbation strengths in subfigures (a) and (b)) and small contributions from nonlinearities when employing the second approach. This suggests that for the MPI-ESM $\chi_\beta^{(O)}(t)$ can be equally well recovered by either of the two approaches. Despite this result, I select the second approach for deriving $\chi_\beta^{(O)}(t)$ in the next sections because this approach accounts for nonlinearities that may cause a more significant error in the recovery for other models. But because subfigure (b) indicates that non-negligible nonlinearities may still remain in the response even when employing this second approach (because of the slightly increasing trend), I choose conservatively to take data only until the optimal forcing strength (138 ppm or 50 years of the 1% bgc experiment) and not the full time series. As

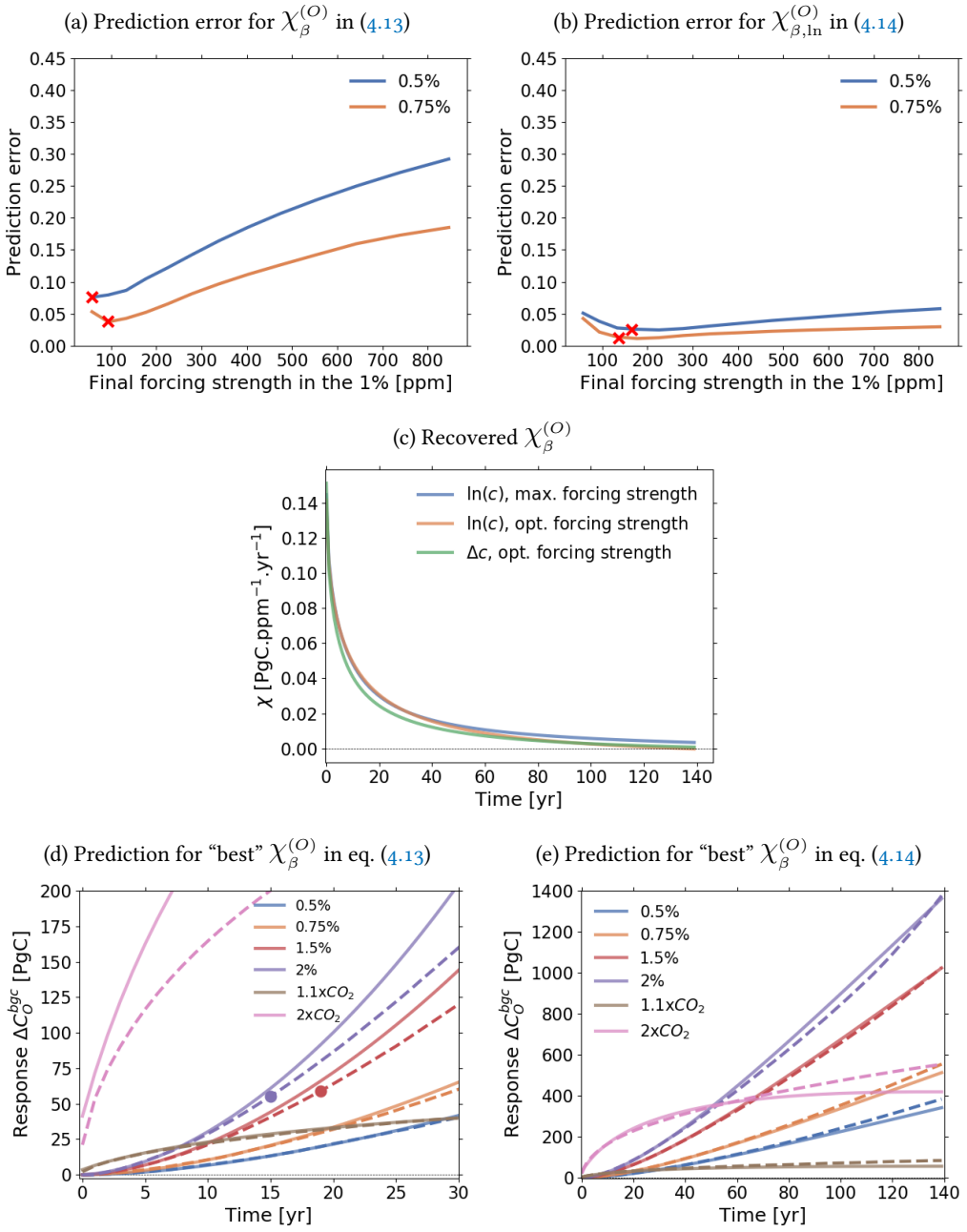


Figure 4.1 : Generalized sensitivity $\chi_\beta^{(O)}(t)$ and prediction of responses from additional experiments. (a) Prediction error (4.15) for $\chi_\beta^{(O)}(t)$ derived with the first approach employing eq. (4.13) for prediction. (b) Prediction error (4.15) for $\chi_\beta^{(O)}(t)$ derived with the second approach employing eq. (4.14) for prediction. (c) Response function $\chi_\beta^{(O)}(t)$ recovered with the first approach at optimal forcing strength (Δc), second approach at optimal forcing strength ($\ln(c)$, opt. forcing strength), and second approach at maximal forcing strength ($\ln(c)$, max. forcing strength). (d) Prediction of additional experiments taking the "best" recovery of $\chi_\beta^{(O)}(t)$ (using the second approach at optimal forcing strength) employed in eq. (4.13). (e) Prediction of additional experiments taking the "best" recovery of $\chi_\beta^{(O)}(t)$ (using the second approach at optimal forcing strength) employed in eq. (4.14). Continuous lines are predictions and dashed lines are responses from the MPI-ESM. Dots indicate the maximum value for which responses are predictable by eq. (4.13) according to the estimate of the linear regime (see text). For better visibility of the regions within the linear regime in (c), the responses are shown only for the first 30 years. For more details see text.

suggested by the small differences between the recoveries for the optimal and maximal forcing strengths in subfigure (c), taking data only until this optimal forcing strength does not significantly hinder the recovery of time scales at the order of the length of the full time series, which is in agreement with the results found for the land carbon in chapter 3.

To obtain evidence that the recovered $\chi_{\beta}^{(O)}(t)$ indeed characterizes the biogeochemical response of the ocean carbon of the MPI-ESM to any temporal development of weak atmospheric CO_2 perturbations, I show how well it predicts the model response for additional CO_2 perturbation experiments (see Table 2.1). The prediction is performed by convoluting the recovery of $\chi_{\beta}^{(O)}(t)$ selected above with the different CO_2 perturbation scenarios first via eq. (4.13), which is the equation employed in the generalized framework (see section 4.2, eq. (4.3)), and then via eq. (4.14), to see if by this equation the prediction is indeed further improved for all experiments. The result for the first case is shown in Fig. 4.1d. To better visualize the linear regime region where the approximation works, I show only the first 30 years. The responses for the 1.5% and 2% bgc experiments are marked with dots where the forcing strength exceeds 94 ppm, the maximal forcing strength corresponding to the linear regime estimated from subfigure (a). As seen, the recovered $\chi_{\beta}^{(O)}(t)$ predicts almost perfectly the response from experiments whose forcing strengths are within the linear regime estimate, i.e. the 0.5%, 0.75% and $1.1\times CO_2$ bgc experiments. It also predicts with reasonable quality of agreement the response from the 1.5% and 2% bgc experiments for forcing strengths within the estimated linear regime. As forcing strengths exceed the linear regime, the prediction starts to deviate from the model response. Likewise, the prediction fails for the whole $2\times CO_2$ bgc response because its forcing strength is larger than the linear regime estimate throughout the whole experiment. These results therefore suggest that indeed $\chi_{\beta}^{(O)}(t)$ characterizes the biogeochemical response of ocean carbon for any temporal development of atmospheric CO_2, as long as its strength is within the linear regime.

Fig. 4.1e shows the prediction employing eq. (4.14). As seen, now the predictive power of $\chi_{\beta}^{(O)}(t)$ improves considerably, extending to the whole time series for all experiments. Some discrepancies are nevertheless encountered, especially for the last years of the $2\times CO_2$ bgc response. Such discrepancies are probably related to the relatively limited amount of information that data from 1%-experiments provide to recover response functions, as observed for the land carbon in chapter 3.

Despite these discrepancies, overall subfigures (d) and (e) show that the recovered $\chi_{\beta}^{(O)}(t)$ predicts the response from different experiments with reasonable quality up to certain perturbation strengths. Eq. (4.14) – employed for the prediction in subfigure (e) – demonstrated to predict the response of the model up to higher perturbation strengths than eq. (4.13). Therefore, eq. (4.14) is probably more appropriate when the aim is to predict the separate biogeochemical response of the model to prescribed atmospheric CO_2 perturbations. On the other hand, eq. (4.13) is the equation employed in the generalized α-β-γ framework (first term on the right-hand side of eq. (4.3)). Therefore, when estimating the linear regime for the application of the generalized α-β-γ framework one has to consider for the ocean bio-

geochemical response the estimate obtained from eq. (4.13), i.e. the linear regime estimate indicated by subfigure (a).

The conclusions from this subsection therefore suggest that the best approach to derive $\chi_\beta^{(O)}(t)$ in the next sections is the second approach (eq. (4.14)), taking data until 138 ppm or 50 years of the 1% bgc experiment, which corresponds to the first minimum for the prediction error in Fig. 4.1b. In addition, the linear regime for the ocean biogeochemical response in the generalized α-β-γ framework (first term in the right-hand side of eq. (4.3)) is of about 94 ppm, as estimated from Fig. 4.1a.

4.3.2 *Generalized sensitivity* $\chi_\gamma^{(O)}$

Approach to recover $\chi_\gamma^{(O)}$

In this subsection I recover $\chi_\gamma^{(O)}(t)$. Since, as will be seen, no strong nonlinearities are present in the response, $\chi_\gamma^{(O)}(t)$ is recovered only from

$$\Delta C_O^{rad}(t) = \int_0^t \chi_\gamma^{(O)}(t-s)\Delta T_O(s)ds, \tag{4.16}$$

where $\Delta C_O^{rad}(t)$ is the ocean carbon response in the "rad" setup. Because the quality of the recovery by the RFI method depends on the quality of the estimate of the noise in the data, and this estimate may be improved by enforcing monotonicity (see section 2.3.5), I assume that $\chi_\gamma^{(O)}$ is monotonic and recover it enforcing monotonicity. Although the arguments from Appendix A.4 to demonstrate monotonicity do not apply to $\chi_\gamma^{(O)}$, Appendix C.1 shows that the recovery from the RFI method enforcing monotonicity approximately matches the recovery from an independent method that does not enforce monotonicity[3], suggesting that monotonicity is a plausible assumption. Further indication of the quality of the recovery will be obtained below by checking how well the recovered $\chi_\gamma^{(O)}$ predicts the model response from different experiments.

Recovery of $\chi_\gamma^{(O)}$

Following the procedure from last section, I firstly evaluate the linear regime of the ocean radiative response. For this purpose I derived $\chi_\gamma^{(O)}(t)$ from the 1% rad experiment and employed it in eq. (4.16) to predict the 0.5% and 0.75% rad responses. The prediction error is plotted in Fig. 4.2a as a function of the final forcing strength in the 1% rad experiment. As seen, the error decreases with increasing final forcing strength, indicating that no strong

3 This method is however not appropriate for our application to CMIP5 models because it requires data from tailored perturbation experiments that are not available in C^{4}MIP databases.

nonlinearities are present in the response. As a result, the response can be considered linear for the whole range of forcing strengths in the 1% rad experiment (temperatures up to around 4 K). Hence I choose to recover $\chi_\gamma^{(O)}(t)$ in the next sections by taking data from the full 1% rad experiment.

The resulting $\chi_\gamma^{(O)}(t)$ recovered in this way is shown in Fig. 4.2b. The negative values reflect the fact that as temperatures rise, globally the ocean loses carbon to the atmosphere. This can be better seen in subfigure (c), which shows the model responses from the different rad experiments and the prediction from the recovered $\chi_\gamma^{(O)}(t)$. Because in these experiments the radiative effect of CO_2 is switched on (while the biogeochemical effect is switched off), temperatures rise, leading to the mentioned global loss of ocean carbon to the atmosphere reflected by the negative responses. The quality of agreement between the model responses and the predictions from $\chi_\gamma^{(O)}(t)$ suggests that indeed the recovered $\chi_\gamma^{(O)}(t)$ characterizes the radiative response of ocean carbon to any temporal development of weak CO_2 perturbations. The larger discrepancy in the prediction over the last years of the $2\times CO_2$ rad response was also observed for the land carbon in section 3.3, and as explained there is probably related to the limited amount of information provided by data taken from the 1% experiment.

In summary, since the response can be considered linear over the whole 1% rad experiment I choose to derive $\chi_\gamma^{(O)}(t)$ in the next subsections from eq. (4.16) taking data from the full experiment.

4.3.3 *Generalized sensitivities $\chi_\alpha^{(L)}$ and $\chi_\alpha^{(O)}$*

Approaches to recover $\chi_\alpha^{(L)}$ and $\chi_\alpha^{(O)}$

Following subsection 4.3.1 I recover $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$ employing two different approaches. In the first approach I recover them from eqs. (4.4) and (4.5) from the generalized framework. From this first approach I will also obtain an estimate of the linear regime for the temperature responses as described in the framework.

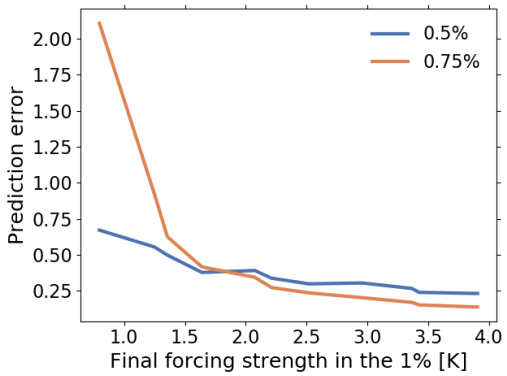
But because CO_2 radiative forcing is known to increase logarithmically with atmospheric CO_2 concentration (Myhre et al., 1998), I derive $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$ also from a second approach that takes this relationship into account. This approach consists in deriving the response functions from

$$\Delta T_L(t) = \int_0^t \chi_{\alpha,\ln}^{(L)}(t-s) c_{PI} \ln\frac{c(s)}{c_{PI}} ds, \tag{4.17}$$

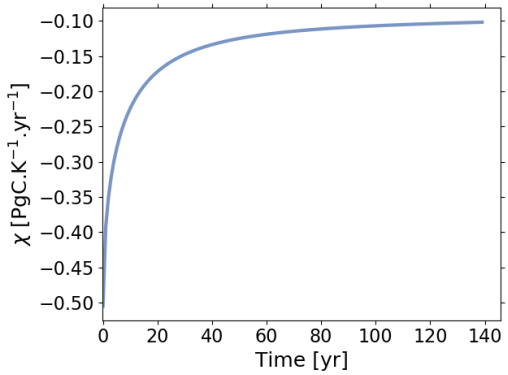
$$\Delta T_O(t) = \int_0^t \chi_{\alpha,\ln}^{(O)}(t-s) c_{PI} \ln\frac{c(s)}{c_{PI}} ds. \tag{4.18}$$

As explained in section 3.4, the resulting $\chi_{\alpha,\ln}^{(L)}(t) = \chi_\alpha^{(L)}(t)$ and $\chi_{\alpha,\ln}^{(O)}(t) = \chi_\alpha^{(O)}(t)$.

(a) Prediction error for $\chi_\gamma^{(O)}$ in (4.16)



(b) Recovered $\chi_\gamma^{(O)}$



(c) Prediction for $\chi_\gamma^{(O)}$ in (4.16)

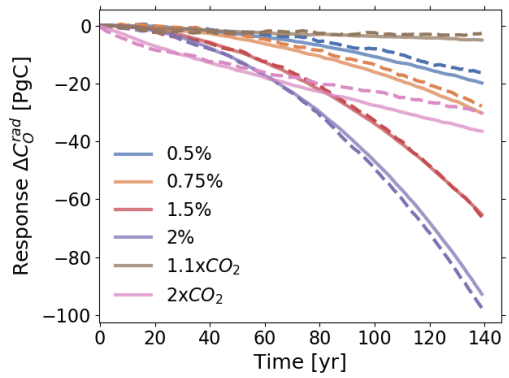


Figure 4.2 : Generalized sensitivity $\chi_\gamma^{(O)}(t)$ and prediction of responses from additional experiments. (a) Prediction error (4.15) employing the recovered $\chi_\gamma^{(O)}(t)$ in eq. (4.16). (b) Recovered response function $\chi_\gamma^{(O)}(t)$. (c) Prediction of additional experiments employing the recovered $\chi_\gamma^{(O)}(t)$ in eq. (4.16). Continuous lines are predictions and dashed lines are responses from the MPI-ESM. For more details see text.

Recovery of $\chi_\alpha^{(L)}$ and $\chi_\alpha^{(O)}$

Following the reasoning from subsection 4.3.1, I first evaluate the linear regime for the temperature responses in the generalized α-β-γ framework, which are described by (4.4) and (4.5). For this purpose I recovered $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$ from the first approach (eqs. (4.4) and (4.5)) taking data from the 1% rad experiment and then employed the recovered response functions once more in (4.4) and (4.5) to predict the responses from the 0.5% and 0.75% rad experiments. The prediction error is plotted in Figs. 4.3a and 4.4a. As seen, in both figures the error decreases until about 279 ppm, presenting after that in Fig. 4.3a a clear increase and in Fig. 4.4a a slight decrease followed by an increase for the 0.5% rad response and an approximately constant behaviour for the 0.75% rad response. Therefore I conservatively estimate for both land and ocean temperature responses the linear regime as perturbation strengths below 279 ppm. But because this range of forcing strengths is larger than for the land and ocean biogeochemical responses (see subsection 3.4 and subsection 4.3.1), it is the linear regime of the biogeochemical responses that limits the applicability of the generalized α-β-γ framework as a whole.

To assess whether the second approach indeed improves the recovery of $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$, one must check whether eqs. (4.17) and (4.18) indeed account for nonlinearities in the response. If this is the case, eqs. (4.17) and (4.18) should predict the 0.5% and 0.75% rad responses better than eqs. (4.4) and (4.5). To evaluate this, I recovered $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$ from eqs. (4.17) and (4.18) taking data once more from the 1% rad experiment, and then employed the recovered response functions in (4.17) and (4.18) to predict the 0.5% and 0.75% rad responses. The resulting prediction error is shown in Figs. 4.3b and 4.4b. As seen in both figures, now the prediction error clearly decreases for increasing final forcing strength, indicating that no strong nonlinearities are present in the response. This clear decreasing trend in the prediction error indicates that the second approach (eqs. (4.17) and (4.18)) is indeed more appropriate to recover $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$.

Therefore in Figs. 4.3c and 4.4c I show the response functions recovered with the second approach. Plotted are recoveries by the RFI method both enforcing and not enforcing monotonicity. As seen, the recoveries obtained without enforcing monotonicity are approximately monotonic so that differences between the two types of recovery are small. Hence in the next sections I choose to derive $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$ employing the second approach and enforcing monotonicity by the RFI method. As already mentioned, the advantage of enforcing monotonicity is that one may improve the noise level estimate used in the method and thereby improve the quality of the recovery.

To obtain evidence that the recovered $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$ indeed characterize the land and ocean temperature responses to any temporal development of weak CO_2 perturbations, I show how well they predict additional experiments. Following subsection 4.3.1 I first show the predictive power of $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$ when employed in the generalized α-β-γ framework (eqs. (4.4) and (4.5)). In Figs. 4.3d and 4.4d I plot the predictions by eqs. (4.4) and (4.5) taking the recoveries selected above (second approach enforcing monotonicity).

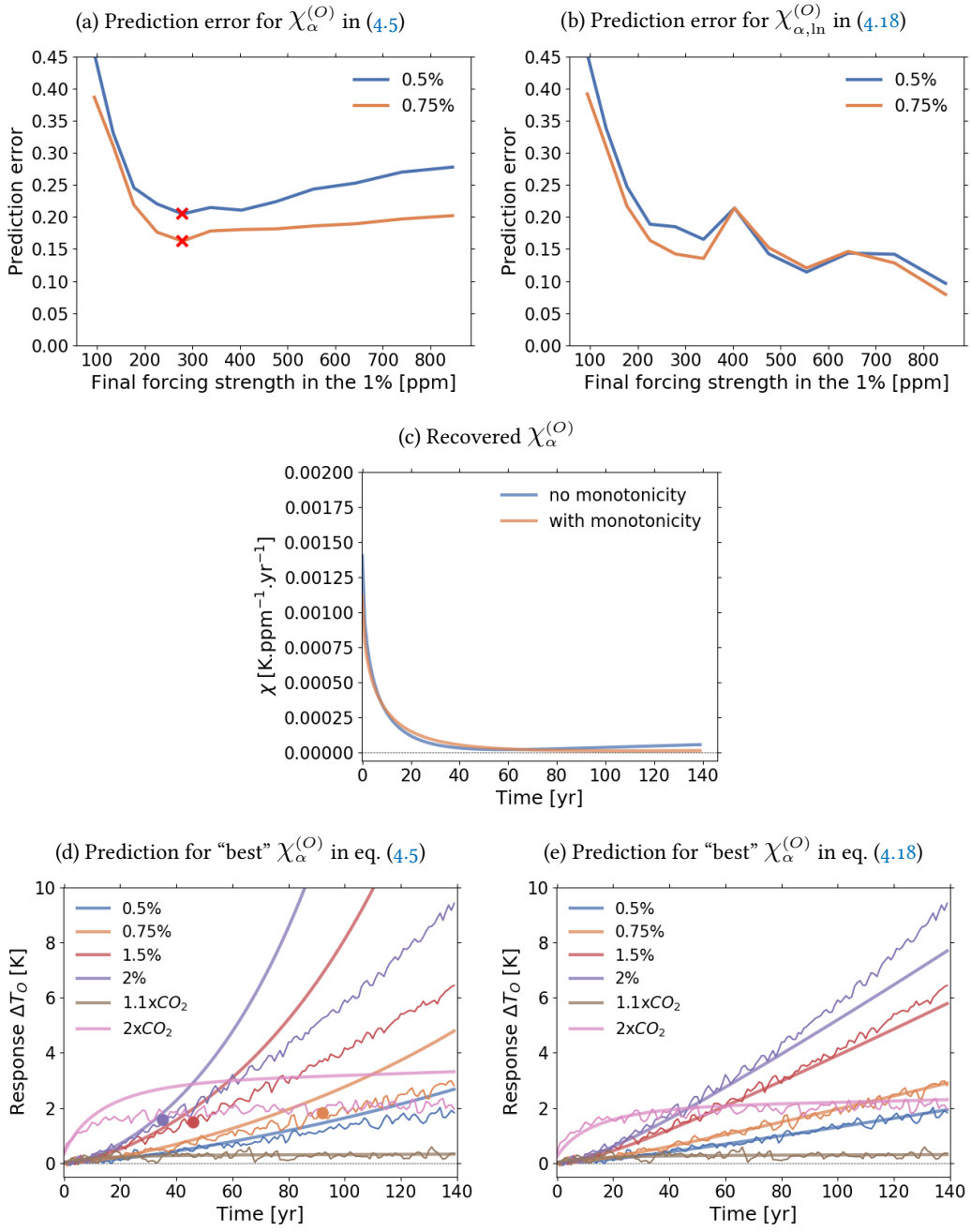


Figure 4.3 : Generalized sensitivity $\chi_\alpha^{(O)}(t)$ and prediction of responses from additional experiments. (a) Prediction error (4.15) for $\chi_\alpha^{(O)}(t)$ derived with the first approach employing eq. (4.5) for prediction. (b) Prediction error (4.15) for $\chi_\alpha^{(O)}(t)$ derived with the second approach employing eq. (4.18) for prediction. (c) Response function $\chi_\alpha^{(O)}(t)$ recovered with the second approach both enforcing and not enforcing monotonicity. (d) Prediction of additional experiments taking the "best" recovery of $\chi_\alpha^{(O)}(t)$ (using the second approach enforcing monotonicity) employed in eq. (4.5). (e) Prediction of additional experiments taking the "best" recovery of $\chi_\alpha^{(O)}(t)$ (using the second approach enforcing monotonicity) employed in eq. (4.18). Thick lines are predictions and thin lines are responses from the MPI-ESM. Dots indicate the maximum value for which responses are predictable by eq. (4.5) according to the estimate of the linear regime. For more details see text.

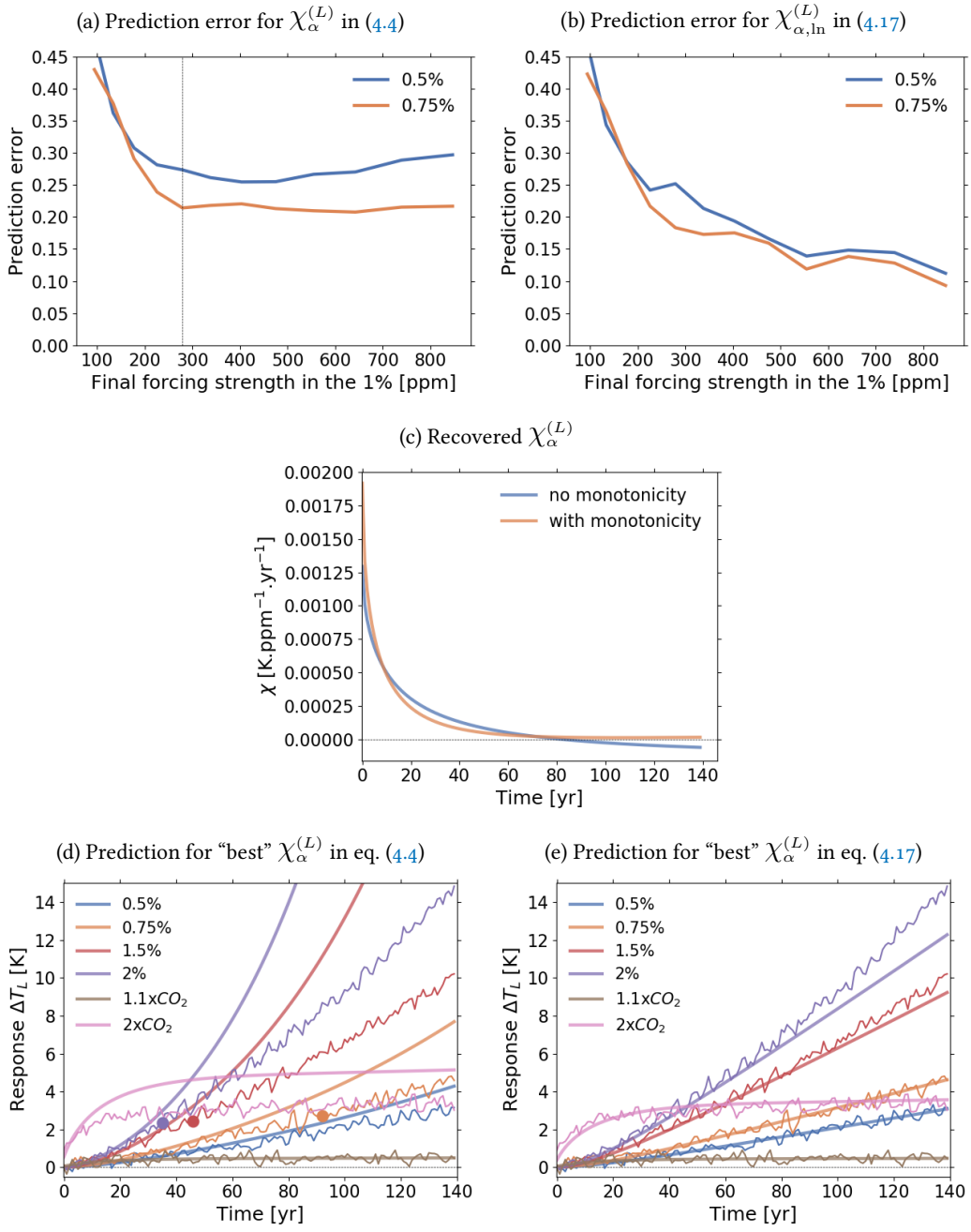


Figure 4.4 : Generalized sensitivity $\chi_\alpha^{(L)}(t)$ and prediction of responses from additional experiments. (a) Prediction error (4.15) for $\chi_\alpha^{(L)}(t)$ derived with the first approach employing eq. (4.4) for prediction. (b) Prediction error (4.15) for $\chi_\alpha^{(L)}(t)$ derived with the second approach employing eq. (4.17) for prediction. (c) Response function $\chi_\alpha^{(L)}(t)$ recovered with the second approach both enforcing and not enforcing monotonicity. (d) Prediction of additional experiments taking the "best" recovery of $\chi_\alpha^{(L)}(t)$ (using the second approach enforcing monotonicity) employed in eq. (4.4). (e) Prediction of additional experiments taking the "best" recovery of $\chi_\alpha^{(L)}(t)$ (using the second approach enforcing monotonicity) employed in eq. (4.17). Thick lines are predictions and thin lines are responses from the MPI-ESM. Dots indicate the maximum value for which responses are predictable by eq. (4.4) according to the estimate of the linear regime. For more details see text.

The responses for the 0.75%, 1.5% and 2% rad experiments are marked with dots where the forcing strength exceeds 279 ppm, the maximal forcing strength corresponding to the linear regime estimated from Figs. 4.3a and 4.4a. As seen, the recovered response functions predict the model responses with some overestimation but still with reasonable quality of agreement[4] for forcing strengths within the estimated linear regime, i.e. for the whole $1.1\times CO_2$ and 0.5% rad experiments, and for the 0.75%, 1.5% and 2% rad experiments for values preceding the dots. For these three latter experiments predictions start to strongly deviate from the responses as forcing strengths exceed the estimated linear regime. Likewise, the predictions fail basically for the whole $2\times CO_2$ rad response because its forcing strength is larger than the linear regime estimate throughout the whole experiment.

As expected from the known logarithmic relationship between radiative forcing and CO_2 concentration, the predictive power of $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$ should improve when employing (4.17) and (4.18) instead of (4.4) and (4.5). To confirm this I plot in Figs. 4.3e and 4.4e the predictions by eqs. (4.17) and (4.18) taking the selected recoveries of $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$. As seen, the predictions indeed improve considerably. Almost all responses can be predicted for the whole time series. The exceptions are the last years of the 1.5% and 2% rad responses, which indicates a deviation from linearity for those levels of perturbation strength.

As in section 4.3.1, overall the prediction results suggest that indeed the recovered response functions characterize the land and ocean temperature responses to any temporal development of atmospheric CO_2 up to certain perturbation strengths. While eqs. (4.17) and (4.18) are more appropriate when the aim is simply to predict model responses because of their extended predictive power, eqs. (4.4) and (4.5) are the equations actually employed in the generalized α-β-γ framework so that for the application of the framework one must consider the linear regime of the temperature responses estimated from Figs. 4.3a and 4.4a.

In summary, the conclusions from this subsection suggest that the best approach to derive $\chi_\alpha^{(L)}(t)$ and $\chi_\alpha^{(O)}(t)$ in the next sections is the second approach (eqs. (4.17) and (4.18)), taking data from the full 1% rad experiment. The linear regime for the land and ocean temperature responses in the generalized α-β-γ framework (eqs. (4.5) and (4.4)) is of about 279 ppm, as estimated from Figs. 4.3a and 4.4a. But as already mentioned, because the linear regime for the biogeochemical response of land and ocean carbon is restricted to forcing strengths even smaller than that for (4.4) and (4.5) – and obviously also smaller than that for the radiative responses of land and ocean carbon, which are linear for the whole 1% rad experiment as seen in sections 3.3 and 4.3.2 –, the applicability of the generalized framework as a whole is limited by the linear regime of the biogeochemical responses, which is estimated as forcing strengths up to about 94 ppm.

4 Note that linear response functions characterize only the ensemble average of the response (see section 2.2) so that they cannot predict high-frequency temperature variations arising from the internal variability of the system.

Selection of approaches to recover $\chi_\gamma^{(L)}(t)$ and $\chi_\beta^{(L)}(t)$ and general summary

To recover $\chi_\gamma^{(L)}(t)$ and $\chi_\beta^{(L)}(t)$ in the next sections I use the conclusions from chapter 3. Because the radiative response of land carbon can be considered linear for the whole range of perturbation strengths in the 1% rad experiment (section 3.3), $\chi_\gamma^{(L)}(t)$ will be derived from eq. (3.1) taking data from the full 1% rad experiment. And because the obtained $\chi_\gamma^{(L)}(t)$ is monotonic (without enforcing monotonicity), I will assume monotonicity when deriving $\chi_\gamma^{(L)}(t)$ for all other models so that all generalized sensitivites will be derived enforcing monotonicity by the RFI method (as mentioned above, this assumption may further improve the quality of the recovery, as explained in section 2.3.5). To derive $\chi_\beta^{(L)}(t)$, conclusions from section 3.4 suggest the third approach analyzed there (eqs. (3.5) and (3.9)), which derives first the response function $\chi_{NPP}(t)$ to perturbations in Net Primary Production (NPP) and then transforms the obtained $\chi_{NPP}(t)$ into the desired $\chi_\beta^{(L)}(t)$. Because the prediction error plot for $\chi_{NPP}(t)$ presents minima (see Fig. 3.5), following the reasoning from subsection 4.3.1 I choose conservatively to take data only until the first minimum, which corresponds to 279 ppm or 70 years of the 1% bgc experiment.

With the recovery of all generalized sensitivities for the MPI-ESM I complete the methods part of the study. The approaches selected in this section will be employed to recover the generalized sensitivities for all CMIP5 models in the next sections. A general summary of the selected approaches is given in Table 4.1.

4.4 AIRBORNE FRACTION IN THE MPI-ESM DETERMINED FROM THE GENERALIZED FRAMEWORK

In the last section I derived for the MPI-ESM the ocean carbon and temperature generalized sensitivities. Together with the land carbon generalized sensitivities derived chapter 3 they give the main ingredient of the generalized α-β-γ framework. As mentioned above, the first aim of this study is to demonstrate that the generalized framework consistently describes the linear dynamics of the global carbon cycle in the MPI-ESM. This demonstration will add confidence that the framework is appropriate to quantify the climate-carbon cycle feedbacks and airborne fraction in the CMIP5 models in section 4.5. The first part of the demonstration was already done in chapter 3 and section 4.3 by showing that from the generalized sensitivities one can predict the separate land and ocean carbon responses to different CO_2 perturbation scenarios, which indicates that these sensitivities can be employed to study the carbon cycle dynamics in concentration-driven simulations. But since the global carbon cycle is a coupled system that responds to CO_2 emissions, to complete the demonstration one must show that this framework describes also the response of atmospheric CO_2 to CO_2 emissions in emission-driven simulations. Because from the first part of the demonstration one knows how to obtain the response of land and ocean carbon to changes in atmospheric CO_2, by knowing the response of atmospheric CO_2 to

Table 4.1 : Approaches selected to derive the generalized sensitivities in sections 4.4 and 4.5 and linear regime of the responses in the generalized framework according to the analyses from chapter 3 and section 4.3. The column "1% time series length" refers to the length of the time series from the 1% experiment that is taken to derive a particular generalized sensitivity. The column "Linear regime" refers to the range of forcing strengths for which the response characterized by a particular generalized sensitivity in the generalized framework can be considered linear (i.e. the linear regime for each response at the right-hand side of eqs. (4.2)-(4.5)). All generalized sensitivities are derived enforcing monotonicity by the RFI algorithm.

Sensitivity	Approach	1% time series length	Linear regime
$\chi_{\beta}^{(L)}$	Third (eqs. (3.5) and (3.9))	70 yrs	94 ppm
$\chi_{\gamma}^{(L)}$	Eq. (3.1)	140 yrs	at least 6 K
$\chi_{\alpha}^{(L)}$	Second (eq. (4.17))	140 yrs	279 ppm
$\chi_{\beta}^{(O)}$	Second (eq. (4.14))	50 yrs	94 ppm
$\chi_{\gamma}^{(O)}$	Eq. (4.16)	140 yrs	at least 4 K
$\chi_{\alpha}^{(O)}$	Second (eq. (4.18))	140 yrs	279 ppm

CO_2 emissions one can determine also the response of the land and ocean carbon to CO_2 emissions. As explained in section 4.2, the response of atmospheric CO_2 to CO_2 emissions is fully determined by the airborne fraction. Hence, to complete the demonstration it suffices to show that the airborne fraction predicted by the generalized framework is a good approximation to the airborne fraction determined directly from emission-driven simulations. This is the purpose of the present section.

4.4.1 *Determining the airborne fraction from emission-driven simulations*

To understand how the airborne fraction can be determined from emission-driven simulations, one must consider how atmospheric CO_2 responds to CO_2 emissions. For weak CO_2 emissions, the response of atmospheric CO_2 can be described within the linear response framework (see section 2.2) by

$$\Delta c(t) = \int_0^t \chi_\zeta(t-s)e(s)ds, \tag{4.19}$$

where $\chi_\zeta(t)$ is the linear response function that characterizes the response. By Laplace-transforming eq. (4.19) one obtains

$$\Delta\widetilde{c}(p) = \widetilde{\chi}_\zeta(p)\widetilde{e}(p). \tag{4.20}$$

Comparing eq. (4.20) with eq. (4.7) gives (as also noted by Enting and Clisby, 2019)

$$\widetilde{A}(p) = p\widetilde{\chi}_\zeta(p). \tag{4.21}$$

Therefore, the airborne fraction $\widetilde{A}(p)$ can be determined from emission-driven simulations in three steps: First, recover $\chi_\zeta(t)$ from eq. (4.19) taking data for $\Delta c(t)$ and $e(t)$. Second, Laplace-transform the recovered $\chi_\zeta(t)$ to obtain $\widetilde{\chi}_\zeta(p)$. Finally, apply eq. (4.21) to determine $\widetilde{A}(p)$ from $\widetilde{\chi}_\zeta(p)$. In the following I obtain the airborne fraction in this way and then compare it to the airborne fraction obtained by (4.8) employing the generalized α-β-γ framework.

As first step to derive the airborne fraction from emission-driven simulations I recovered $\chi_\zeta(t)$ taking data from an impulse-emission experiment. In this experiment, starting from a pre-industrial control run (see "esmControl" in Taylor et al., 2012), the system is perturbed by a small impulse in emissions of 100 PgC/yr (as in Joos et al., 2013) in the first year. The advantage of such type of experiment is that the impulse response gives approximately also the desired linear response function $\chi_\zeta(t)$ so that errors from the ill-posedness of eq. (4.19) – see chapter 2 – can be avoided. Nevertheless, the disadvantage is that small perturbation strengths lead to a response with poor signal-to-noise ratio, which deteriorates the recovery of $\chi_\zeta(t)$. Therefore instead of taking data from only one realization of the experiment I performed an ensemble of 5 realizations with initial conditions taken 100 years separated from one another (as in Lembo et al., 2020) and took data from the ensemble average of

the response. This procedure is in agreement with linear response theory because strictly the linear response function characterizes only the ensemble average of the response (see section 2.2). With the data from the ensemble average I employed the RFI method to recover $\chi_\zeta(t)$. Although in principle a special method is not needed to recover $\chi_\zeta(t)$ because the impulse response gives also the linear response function, by employing the RFI method I obtain not only $\chi_\zeta(t)$ but also the spectrum of the response (see section 2.3) from which the Laplace transform $\widetilde{\chi}_\zeta(p)$ in eq. (4.21) can be analytically calculated.

The resulting impulse response after taking the ensemble average and the fit by the recovered $\chi_\zeta(t)$ are shown in Fig. 4.5a. One sees that the experiment is not a perfect impulse experiment because the impulse extends even beyond the first year. This may be related to internal interpolations in the model when computing the emissions within the time interval of one year. Such problem leads to an error in the estimation of $\chi_\zeta(0)$, which must be 1 by conservation of mass: For an impulse in emissions $e(t) = a\delta(t)$ it must be that at $t = 0$ the impulse response $\Delta c_\delta(0) = a\chi_\zeta(0) = a$, hence $\chi_\zeta(0) = 1$. To avoid this error, I recalculated $\chi_\zeta(t)$ using the same regularization parameter obtained from the RFI method but employing a Lagrange multiplier to account for this constraint (see Appendix C.2). The result for both recoveries is shown in Fig. 4.5b. The response functions are almost identical except for $\chi_\zeta(0)$, which is corrected by enforcing the constraint $\chi_\zeta(0) = 1$.

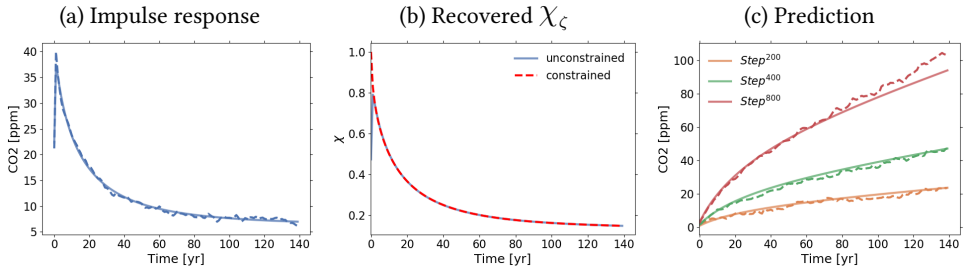


Figure 4.5 : Response function $\chi_\zeta(t)$ recovered from impulse-emission experiment. (a) Impulse response and fit by $\chi_\zeta(t)$. (b) $\chi_\zeta(t)$ recovered with and without enforcing the constraint $\chi_\zeta(0) = 1$. (c) Prediction of different experiments employing the recovered $\chi_\zeta(t)$ in eq. (4.19). For more details see text.

To make sure that the impulse response is within the linear regime and therefore that the recovery of $\chi_\zeta(t)$ is not spoiled by nonlinearities, in line with the procedure in section 4.3 I check the quality of the recovered $\chi_\zeta(t)$ by employing it to predict additional emission experiments via eq. (4.19). For the additional experiments I chose step-emission experiments where starting from the control run the system is perturbed by abrupt, constant emissions of 200, 400 and 800 PgC/yr. The quality of the prediction is shown in Fig. 4.5c. As seen, the obtained $\chi_\zeta(t)$ can predict almost all responses with reasonable accuracy for the whole time series. The only exception is over the last years of the experiment with emissions of 800 PgC/yr. But the discrepancy encountered there is in agreement with results from section 3.4 and section 4.3 that show that strong nonlinear contributions to the biogeochemical response of land and ocean carbon start to arise at about 94 ppm. As a consequence, the

identified discrepancy is not a result of nonlinearities spoiling the recovery of $\chi_\zeta(t)$ but rather a result of nonlinearities that arise in the response of the 800 PgC/yr experiment, which can therefore not be completely predicted by $\chi_\zeta(t)$.

Hence, overall these results suggest that the recovery of $\chi_\zeta(t)$ is not spoiled by nonlinearities and is thus a good candidate to be used to compute the airborne fraction. Therefore $\chi_\zeta(t)$ was Laplace-transformed and the airborne fraction was finally obtained by applying eq. (4.21).

4.4.2 *Airborne fraction determined from the generalized framework*

To obtain the airborne fraction from the generalized framework I took the generalized sensitivities computed with the approaches described in Table 4.1, calculated the Laplace transforms analytically from the recovered spectra of the responses (see section 2.3), and employed the Laplace-transformed sensitivities in eq. (4.8).

Fig. 4.6 compares the results. As seen, the airborne fraction obtained from the generalized framework follows closely that obtained from the emission-driven simulations. Both approaches result in a time-scale dependent airborne fraction that decreases as the time scale increases, starting at 1 and reaching about 0.26 at $1/p$ = 100 years. This time-scale dependence arises because here the airborne fraction is defined differently from the typical approach (e.g. Canadell et al., 2007, Raupach et al., 2008, Gregory et al., 2009, Gloor et al., 2010, Jones et al., 2013): As shown by eq. (4.7) the airborne fraction $\widetilde{A}(p)$ is not a ratio of atmospheric CO_2 fluxes to CO_2 emission fluxes at a particular time but is instead a function that determines the response of atmospheric CO_2 to any temporal development of weak CO_2 emissions accounting for the effect of climate-carbon cycle feedbacks. From eq. (4.7) one notes also that $\widetilde{A}(p) = p\Delta\widetilde{c}(p)/\widetilde{e}(p) = \Delta\widetilde{c}(p)/(\widetilde{e}(p)/p)$ so that in this framework the same airborne fraction is obtained whether one considers instantaneous or cumulated fluxes[5], which might not be the case in the typical definition (compare Fig. 6 in Gregory et al., 2009).

The observed behaviour of the airborne fraction agrees with theoretical expectations. From the initial value theorem (Beerends et al., 2003, p. 292) and the constraint $\chi_\zeta(0) = 1$ it follows that

$$\lim_{p\to\infty} \widetilde{A}(p) \overset{(4.21)}{=} \lim_{p\to\infty} p\widetilde{\chi}_\zeta(p) = \lim_{t\to 0^+} \chi_\zeta(t) = 1, \tag{4.22}$$

5 This conclusion follows by individually inverting the Laplace transforms involved in the definition of $\widetilde{A}(p)$ (see e.g. Doetsch, 1974, p. 317). Written as $\widetilde{A}(p) = p\Delta\widetilde{c}(p)/\widetilde{e}(p)$, the airborne fraction gives the relation between instantaneous fluxes of atmospheric CO_2 and CO_2 emissions, since $\mathcal{L}^{-1}\{\widetilde{e}(p)\} = e(t)$ and $\mathcal{L}^{-1}\{p\Delta\widetilde{c}(p)\} = d\Delta c/dt = dc/dt$ for $\lim_{t\to 0^+} \Delta c(t) = 0$. On the other hand, written as $\widetilde{A}(p) = \Delta\widetilde{c}(p)/(\widetilde{e}(p)/p)$ the airborne fraction gives the relation between cumulated fluxes of atmospheric CO_2 and CO_2 emissions, since $\mathcal{L}^{-1}\{\Delta\widetilde{c}(p)\} = \Delta c(t)$ and $\mathcal{L}^{-1}\{\widetilde{e}(p)/p\} = \int_0^t e(s)ds$.

so that indeed $\widetilde{A}(p)$ must approach 1 as $1/p \to 0$. On the other hand, using the final value theorem (Beerends et al., 2003, p. 293) and assuming that $\chi_\zeta(t)$ approaches zero as $t \to \infty$ one obtains

$$\lim_{p\to 0} \widetilde{A}(p) \overset{(4.21)}{=} \lim_{p\to 0} p\widetilde{\chi}_\zeta(p) = \lim_{t\to\infty} \chi_\zeta(t) = 0, \tag{4.23}$$

so that $\widetilde{A}(p)$ must approach zero as $1/p \to \infty$. In addition, if $\chi_\zeta(t)$ is indeed non-negative and monotonic for all times t as suggested by the results in Fig. 4.5, then $0 \leqq \widetilde{A}(p) \leq 1$ for all time scales $1/p$ (see Appendix C.3). More insight on the behaviour of $\widetilde{A}(p)$ will be given in section 4.5 when discussing its dependence on the climate-carbon cycle feedbacks calculated for the CMIP5 models.

The larger discrepancy between the two estimates of the airborne fraction observed at time scales smaller than 7 years is expected from two types of error that might have affected the results. The first type affects the airborne fraction determined from emission-driven simulations and arises from the fact that $\chi_\zeta(t)$ was not derived from a perfect impulse experiment. Although the constraint $\chi_\zeta(0) = 1$ was enforced, the recovery of $\chi_\zeta(t)$ at small time scales might still not be fully correct. The second type of error affects the airborne fraction derived from the generalized framework and arises from the ill-posedness of (4.12) that must be solved to derive the generalized sensitivities employed in eq. (4.8). The ill-posedness obscures information at small time scales and therefore deteriorates the recovery of the sensitivities at those scales (see chapter 2). Despite these discrepancies, as time scales approach $1/p$ = 0.01 the agreement improves once more, in line with the expectations discussed above.

Overall, the close agreement between the two estimates of the airborne fraction shows that indeed the generalized α-β-γ framework describes with reasonable approximation the linear dynamics of the global carbon cycle in the MPI-ESM even in emission-driven simulations, where climate and the different subsystems of the carbon cycle are coupled. In addition, the agreement of the obtained $\widetilde{A}(p)$ with theoretical expectations adds confidence that the numerical methods employed here can be trusted. These two results suggest that the generalized framework as well as our numerical methods are appropriate to quantify the climate-carbon cycle feedbacks and the airborne fraction in the CMIP5 models in next section.

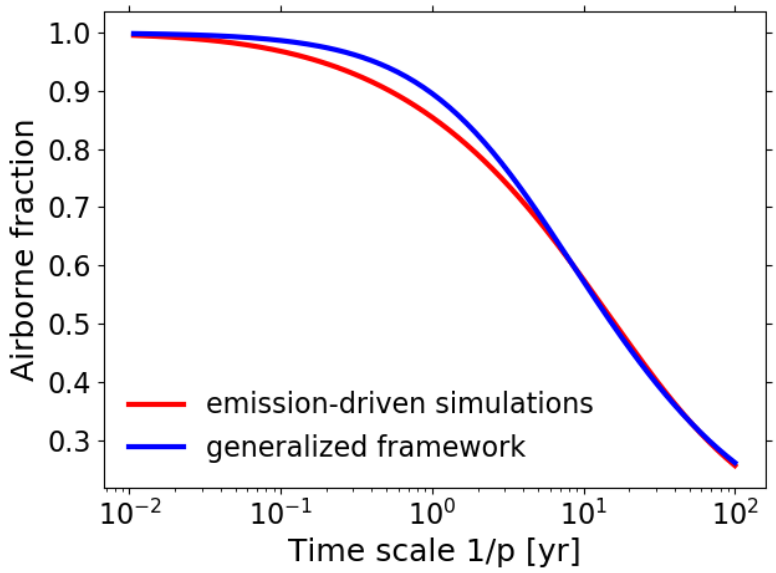


Figure 4.6 : Quality of agreement between the airborne fraction determined from emission-driven simulations and that determined from the generalized α-β-γ framework. The close agreement shows that generalized α-β-γ framework gives a reasonable description of the linear dynamics of the coupled global carbon cycle in the MPI-ESM.

4.5 CLIMATE-CARBON CYCLE FEEDBACKS FOR WEAK PERTURBATIONS IN CMIP5 MODELS

So far, it has been demonstrated that the generalized α-β-γ framework gives a reasonable description of the dynamics of the global carbon cycle under weak atmospheric CO_2 perturbations in a particular Earth System Model. This section is dedicated to the second goal of this study, namely to quantify by this framework the climate-carbon cycle feedbacks and the airborne fraction in different CMIP5 models. As discussed above, the motivation for employing the generalized α-β-γ framework instead of Friedlingstein's framework to quantify the feedbacks is that, in contrast to Friedlingstein's framework, the generalized framework gives a characterization of the feedbacks that is valid for any CO_2 perturbation scenario. Further, by this generalization the strength of the feedbacks is given at each time scale so that the framework gives information about the internal memory structure of the system. But since the generalized framework gives a description of the dynamics of the global carbon cycle, it may be applied not only to quantify the feedbacks but also to study other problems. As shown in last section, another application is to quantify at different time scales the airborne fraction in a way that is independent of the perturbation scenario. Because the airborne fraction is determined from the climate-carbon cycle feedbacks, additionally one can analyze how the different feedbacks influence the airborne fraction. All this will be investigated in this section.

But as discussed in section 4.2, to obtain the feedbacks and the airborne fraction one must first derive the generalized sensitivities. These sensitivities are derived here by the RFI method taking published data from the 1% bgc and 1% rad simulations (Table 2.1) and assuming that the best approaches selected in section 4.3 in the application to the MPI-ESM (Table 4.1) are also appropriate for the application to other CMIP5 models. From the generalized sensitivities the feedbacks are quantified by the feedback functions defined by eqs. (4.10) and (4.11) and the airborne fraction is determined from eq. (4.8).

4.5.1 *Generalized sensitivities*

In this subsection the generalized sensitivities for the CMIP5 models are derived. Although these results depend on the quality of the data (chapter 2) and on how appropriate the numerical approaches selected for the MPI-ESM are for other CMIP5 models, their consistency can be at least qualitatively checked against results obtained by Arora et al., 2013, who also analyzed climate-carbon cycle feedbacks in CMIP5 models. The main difference between their and the present analysis is that in Arora et al., 2013 the sensitivities are quantified by two variations of Friedlingstein's framework (one considering instantaneous and another considering cumulated carbon fluxes), whereas here they are quantified by the generalized framework. Therefore, the sensitivities obtained there (see their Figs. 4-6) depend on the perturbation scenario used for the calculations. In contrast, here I obtain unique generalized sensitivities that characterize the model response in principle for any CO_2 perturbation scenario and for all time scales covered by data, as long as perturbations are sufficiently weak. Thus the consistency check will not be focused on the sensitivities obtained by Arora et al., 2013 but rather on the model responses shown in their Fig. 2, which are independent of their feedback analysis and provide a more direct basis for qualitatively comparing the present results.

The derived sensitivities are shown in Fig. 4.7. In the first row are the sensitivities for the ocean and in the second row the sensitivities for the land. Subfigures (a) and (d) in the first column show the results for the $\chi_\beta^{(O)}(t)$ and $\chi_\beta^{(L)}(t)$ sensitivities that characterize the biogeochemical response of ocean and land carbon. The vertical lines determine the time series length that was taken to derive the sensitivities according to the approaches in Table 4.1. As seen, the sensitivities for almost all models present positive values for all times analyzed. This is because an increase in atmospheric CO_2 concentrations results in an increase in land and ocean carbon stocks: In the land such positive response is a consequence of the CO_2 fertilization effect, which increases plant productivity and vegetation coverage, while in the ocean it is a consequence of the increase in the difference between atmospheric and oceanic CO_2 partial pressure, leading to a positive input flux of CO_2 into the ocean (Arora et al., 2013, Friedlingstein et al., 2006). The surprising negative values of $\chi_\beta^{(L)}(t)$ for the HadGEM2-ES after 70 years are most likely a consequence of nonlinearities in the response of this model. The reason is that $\chi_\beta^{(L)}(t)$ should decrease monotonically to zero (see Appendix A.4), so that in principle it could not get negative unless noise or nonlinearities

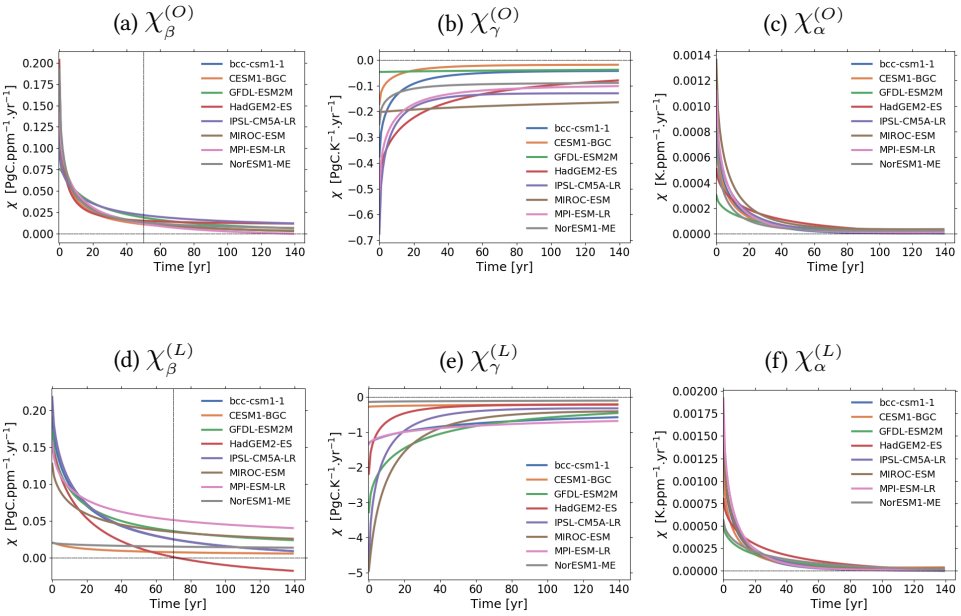


Figure 4.7 : Generalized sensitivities in CMIP5 models.

are deteriorating its recovery (see section 2.4). Because the order of magnitude of the estimated signal-to-noise ratio in the HadGEM2-ES response is equal or larger than that in the response of the other models (not shown), probably the negative values are thus related to nonlinearities. In fact, the monotonicity constraint was accounted for by the RFI method, but because the time series used for the recovery had only 70 years (see Table 4.1), the recovery turned out to be monotonic only within this time range. This result suggests that for this model probably smaller perturbation strengths should be taken to derive $\chi_\beta^{(L)}(t)$ without the influence of strong nonlinearities.

In line with the close model agreement found by Arora et al., 2013 for the ocean biogeochemical responses (compare their Fig. 2i), there is a close agreement between the obtained generalized sensitivities $\chi_\beta^{(O)}(t)$ in subfigure (a). In most models $\chi_\beta^{(O)}(t)$ decays rapidly in the first years and slowly afterwards, reaching small values at large times.

Such overall agreement is in contrast with the results for $\chi_\beta^{(L)}(t)$ in subfigure (d), which, also in line with results from Arora et al., 2013, presents a large model spread. Particularly different are the $\chi_\beta^{(L)}(t)$ sensitivities for the NorESM1-ME and CESM1-BGC models, which have very small values for all times. This behaviour agrees with the overall small biogeochemical response of these models (Fig. 2f of Arora et al., 2013), and can be explained by the fact that these models account for the coupling between nitrogen and carbon cycle, which reduces the CO_2 fertilization effect by nitrogen limitation (Zaehle et al., 2010, Arora et al., 2013). Excluding NorESM1-ME and CESM1-BGC, visually models agree better at small than at large times.

Subfigures (b) and (e) present the results for the $\chi_\gamma^{(O)}(t)$ and $\chi_\gamma^{(L)}(t)$ sensitivites that characterize the radiative response of ocean and land carbon. In agreement with the model responses shown in Arora et al., 2013 (see their Figs. 2e and 2h), both sensitivities are negative for all times. This is because globally ocean and land lose carbon to the atmosphere when only the radiative effect of CO_2 is taken into account. As explained by Arora et al., 2013, this loss is associated to the increase in temperatures that over land results in an increase in heterotrophic respiration and a decrease in globally averaged Net Primary Production, and over ocean results in a decrease in CO_2 solubility.

In contrast to $\chi_\beta^{(O)}(t)$, $\chi_\gamma^{(O)}(t)$ shows a visually large model spread over the whole time series (compare subfigures (a) and (b)). Particularly different $\chi_\gamma^{(O)}(t)$ sensitivities are found for the MIROC-ESM and GFDL-ESM2M models: While for all other models this sensitivity has a rapid decrease in magnitude in the first years, for these models only a slow decrease is observed. In fact the decrease is so slow that it looks as if they had constant values throughout the whole time series, which is only an impression due to the scale of the plot (not shown). While the very small values of GFDL-ESM2M cannot be compared to results from Arora et al., 2013 because this model was not assessed there, the large values of MIROC-ESM are in agreement with its relatively large radiative ocean response (compare Figs. 2h, 5c and 6f in Arora et al., 2013).

In agreement with the difference in magnitude between the radiative responses of land and ocean carbon found in Arora et al., 2013 (compare their Figs. 2e and 2h), $\chi_\gamma^{(L)}(t)$ is at least at small times much larger than $\chi_\gamma^{(O)}(t)$ for almost all models except for NorESM1-ME and CESM1-BGC. As explained by Arora et al., 2013, the coupling in these models between nitrogen and carbon cycle not only leads to a small biogeochemical response but also to a small radiative response because nitrogen remineralization enhances plant productivity, which counteracts the carbon loss from the enhanced soil respiration that is caused by higher temperatures (see also Melillo et al., 2002, Thornton et al., 2009). As in the case of $\chi_\gamma^{(O)}(t)$ in MIROC-ESM and GFDL-ESM2M, $\chi_\gamma^{(L)}(t)$ in NorESM1-ME and CESM1-BGC decays so slowly that because of the scale it seems that they have a constant value throughout the time series, which is also only an impression due to the scale (not shown). Overall there is a better model agreement for $\chi_\gamma^{(L)}(t)$ at large rather than at small times.

Subfigures (c) and (f) present the results for the $\chi_\alpha^{(O)}(t)$ and $\chi_\alpha^{(L)}(t)$ sensitivites that characterize the response of ocean and land temperature to atmospheric CO_2 perturbations. In both cases there is visually a relatively good agreement among models apart from values at small times, for which the recovery is anyway less robust due to ill-posedness of (4.12) that must be solved to recover the sensitivities (see chapter 2).

Overall, results indicate a visually better model agreement for $\chi_\beta^{(O)}(t)$, $\chi_\alpha^{(O)}(t)$ and $\chi_\alpha^{(L)}(t)$, whereas $\chi_\beta^{(L)}(t)$, $\chi_\gamma^{(O)}(t)$, and $\chi_\gamma^{(L)}(t)$ have larger model spreads. Nevertheless, the magnitude and model spread of the different climate-carbon cycle feedbacks are better comparable by considering not the generalized sensitivities but the feedback functions

$\widetilde{f}_{\beta}^{(L)}(p)$, $\widetilde{f}_{\gamma\alpha}^{(L)}(p)$, $\widetilde{f}_{\beta}^{(O)}(p)$ and $\widetilde{f}_{\gamma\alpha}^{(O)}(p)$. The reason is that these functions quantify in a dimensionless way the biogeochemical feedback, which is determined only from $\chi_{\beta}(t)$, and the radiative feedback, which is determined from a combination of $\chi_{\gamma}(t)$ and $\chi_{\alpha}(t)$ (see eqs. (4.10) and (4.11)).

4.5.2 *Climate-carbon cycle feedbacks and airborne fraction*

In the following I quantify by the feedback functions the climate-carbon cycle feedbacks and airborne fraction in the different CMIP5 models, and investigate how those feedbacks influence the values of the airborne fraction at different time scales.

In preparation for this analysis, one must realize that the accuracy of the results in this subsection depends on two aspects: 1) the quality of the generalized sensitivities that were recovered in the last subsection by our numerical methods; and 2) the assumption from the generalized framework that under weak perturbations the carbon response to CO_2 can be completely determined by eqs. (4.2) and (4.3), that describe the additive combination of the biogeochemical and radiative response, and by eqs. (4.4) and (4.5), that describe the land and ocean temperature response to CO_2. Ideally, for each model one should fully check these two aspects with the aid of additional simulations, as was done for the MPI-ESM in the last sections. Unfortunately, such additional simulations are at present not available for other models so that a full check is not possible. Nevertheless, since all CMIP5 models provide in addition to the 1% rad and bgc simulations also a 1% "fully-coupled" simulation (a 1% simulation where both the biogeochemical and the radiative effects of CO_2 are active), one can at least check whether the biogeochemical and radiative responses are indeed additive for a certain range of perturbation strengths. To check this additivity, I plot in Fig. 4.8 for each of these models the response for land, ocean and global (land plus ocean) carbon from the 1% fully-coupled experiment along with the sum of the responses from the 1% bgc and 1% rad experiments. If additivity holds for a certain range of perturbation strengths, then within this range these two curves (1% fully-coupled and the sum of 1% bgc and 1% rad) must agree. As seen, for all models there is indeed a reasonable agreement at least within the linear regime estimate of 94 ppm (section 4.3) for all land, ocean and global carbon, albeit somewhat larger discrepancies are observed for the land and global carbon in the MIROC-ESM and NorESM1-ME. Hence, within the linear regime one may say that the biogeochemical and radiative response are approximately additive in the CMIP5 models. This result, together with the qualitative consistency of the obtained generalized sensitivities when compared to the results by Arora et al., 2013, gives some confidence that our numerical methods and the framework are describing with reasonable approximation the linear dynamics of the carbon cycle in these models.

Therefore, one may proceed to estimate the feedbacks and the airborne fraction for each model via eqs. (4.10), (4.11) and (4.8). The resulting estimates are shown in Fig. 4.9. Starting with the airborne fraction in subfigure (a), one sees that almost all CMIP5 models present a time-scale dependent airborne fraction that decreases as the time scale increases, starting all

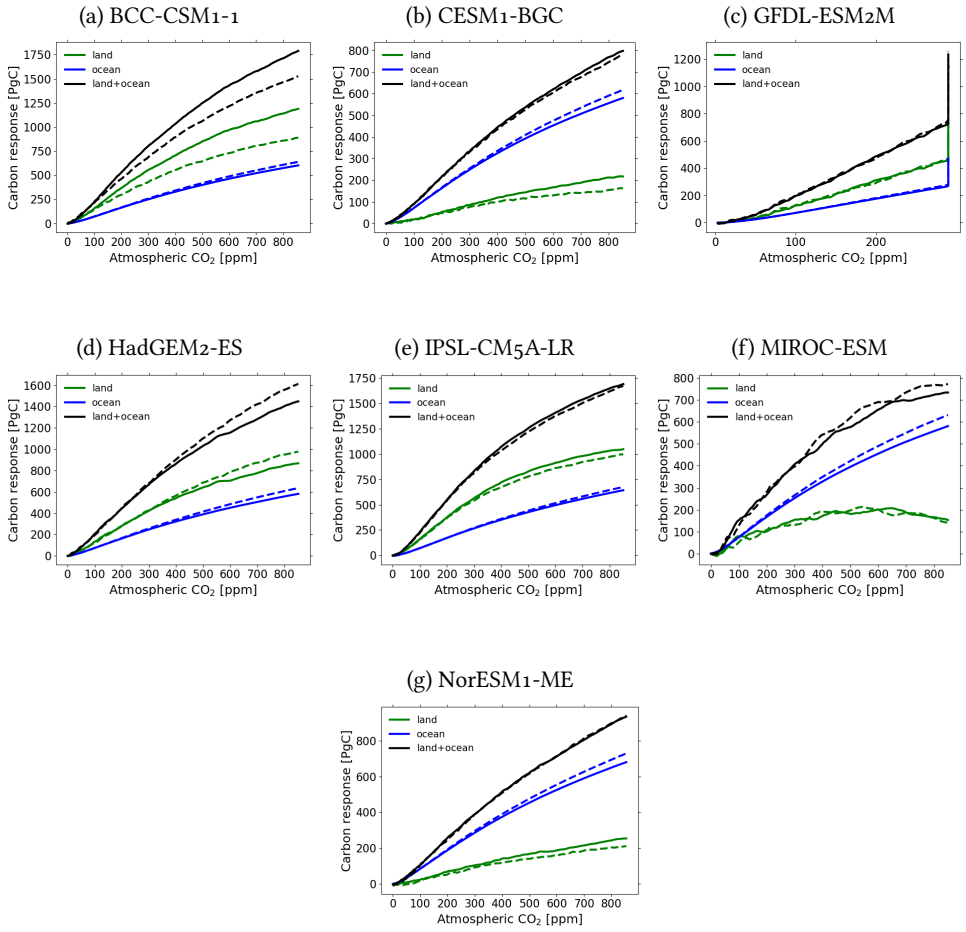


Figure 4.8 : Check of the additivity of the biogeochemical and radiative carbon responses in CMIP5 models. Plotted are the sum of the responses from the 1% bgc and 1% rad experiments (dashed lines) against the response from the 1% fully-coupled experiment (continuous lines) for land, ocean, and global carbon (land plus ocean). Additivity holds approximately within the estimated linear regime of 94 ppm for all models. For more details see text.

at 1 for small time scales and spreading to a range of values from 0.56 to 0.75 at a time scale of 10 years, and from 0.26 to 0.5 at a time scale of 100 years. The exception is the HadGEM2-ES, that presents also a time-scale dependent airborne fraction but that slightly increases at long time scales, which as will be seen is related to a reduction in the magnitude of its land biogeochemical feedback. As explained in section 4.4, the behaviour of the airborne fraction starting at 1 and approaching zero for long time scales is theoretically expected considering that the response function $\chi_\zeta(t)$ has 1 as initial value and approaches zero for long times (see eqs. (4.22) and (4.23)). But the physical meaning of $\widetilde{A}(p)$ may be more intuitively understood by considering the impulse response of atmospheric CO_2 flux. In an impulse-emission experiment, the response of atmospheric CO_2 is simply $\Delta c_\delta(t) = \chi_\zeta(t)$ (see section 4.4). As a result, the atmospheric CO_2 flux can symbolically be written as

$$c'_\delta(t) = \delta(t) + \chi'_\zeta(t), \tag{4.24}$$

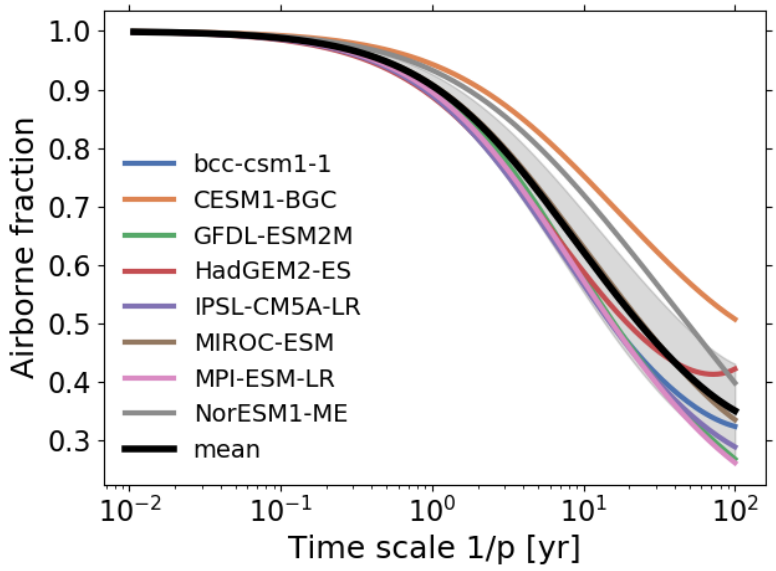
where the prime indicates the first time derivative, c'_δ is the impulse response of atmospheric CO_2 flux (considered as distribution; see Appendix C.3), and $\Delta c'(t) = c'(t)$ was used considering the pre-industrial equilibrium value of CO_2 constant. The Dirac delta distribution δ arises because $\chi_\zeta(t)$ has a "jump" at $t = 0$, as explained in Appendix C.3. From eq. (C.19) in Appendix C.3 one sees that the right-hand side of eq. (4.24) is simply $\widetilde{A}(p)$ transformed to the time domain, since $\widetilde{A}(p)$ can be written as

$$\widetilde{A}(p) = 1 + \widetilde{\chi'_\zeta}(p) = \widetilde{c'_\delta}(p). \tag{4.25}$$

Therefore, one can interpret the airborne fraction studied here as the impulse response of atmospheric CO_2 flux to the CO_2 emissions flux. In fact, by entering the first equality of eq. (4.25) into eq. (4.7) and using the differentiation property of the Laplace transform of distributions so that $p\Delta\widetilde{c}(p) = \Delta\widetilde{c'}(p)$ (Beerends et al., 2003, Table 10), one obtains the same equation derived by Enting, 2007 to describe the response of atmospheric CO_2 flux to the CO_2 emissions flux (see his eq. (7a)), but without the need to assume $\lim_{t\to 0^+} \Delta c(t) = 0$ – also assumed in the footnote of section 4.4.2 –, which does not hold in an impulse experiment because $\Delta c_\delta(0) = \chi_\zeta(0) = 1$. With this interpretation of the airborne fraction one can intuitively understand the results in subfigure (a): In an impulse-emission experiment, at $t = 0$ the atmospheric CO_2 flux is simply equal to the emissions impulse $\delta(t)$; correspondingly, as the time scale $1/p \to 0$ the airborne fraction $\widetilde{A}(p) \to 1$. In contrast, for large times, as atmospheric CO_2 slowly returns to its equilibrium value, the atmospheric CO_2 flux approaches zero; correspondingly, for large time scales $1/p$ also the airborne fraction $\widetilde{A}(p) \to 0$.

How the airborne fraction changes in the time-scale domain is determined by the climate-carbon cycle feedbacks. Indeed, by eq. (4.8) it is seen that the time-scale dependence of $\widetilde{A}(p)$ arises from the time-scale dependence of the net feedback function $\widetilde{f}(p)$. This is because the memory observed in atmospheric CO_2 arises from the processes developing at different time scales in the land and ocean carbon cycle, which act as feedbacks taking carbon into or out of the atmosphere. As seen in subfigure (b), for $1/p \to 0$ all feedback functions approach zero so that by eq. (4.8) the airborne fraction $\widetilde{A}(p) \to 1$. This can be intuitively

(a) Airborne fraction



(b) Feedback functions

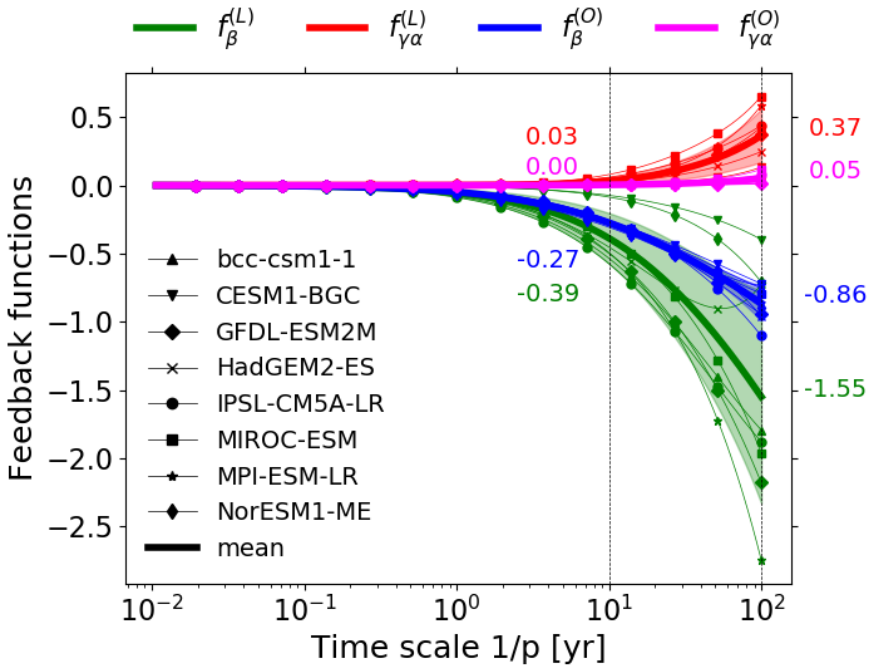


Figure 4.9 : Airborne fraction and climate-carbon cycle feedbacks in CMIP5 models. The numbers in subfigure (b) indicate the model mean for each feedback function at 10-years and 100-years time scales.

understood from the impulse response interpretation of $\widetilde{A}(p)$: At the moment of an impulse in emissions, the land and ocean have not yet started taking up/releasing carbon from/to the atmosphere so that the airborne fraction simply reflects the impulse δ (which is 1 in the time-scale domain). To understand how the airborne fraction behaves as the time scale increases, one must look at the behaviour of the separate feedback functions. For larger time scales, the feedback functions $\widetilde{f}_{\gamma\alpha}^{(L)}$ and $\widetilde{f}_{\gamma\alpha}^{(O)}$ that quantify the radiative feedbacks get increasingly positive, while the feedback functions $\widetilde{f}_{\beta}^{(L)}$ and $\widetilde{f}_{\beta}^{(O)}$ that quantify the biogeochemical feedbacks get in general increasingly negative. The sign of these feedbacks is in agreement with current process understanding (Friedlingstein et al., 2006, Gregory et al., 2009, Arora et al., 2013). But that these feedbacks are either positive or negative for *all* time scales is a non-trivial result, which is explained by the fact that almost all generalized sensitivities in Fig. 4.7 are either always negative or always positive: Since the feedback functions are simply a scaling of the Laplace transform of χ_β and of the product of the Laplace transforms of χ_γ and χ_α (see eqs. (4.10) and (4.11)), along the lines of the argument for $\widetilde{A}(p) \geq 0$ in Appendix C.3 one sees that from the positivity or negativity of these response functions one can determine the sign their Laplace transforms and thereby the sign of the feedback functions for all time scales[6]. Interestingly, from the time scale at which the magnitude of the feedbacks starts to increase, in almost all models – except for the HadGEM2-ES at long time scales – the sum of the biogeochemical feedbacks gets increasingly larger than the sum of the radiative feedbacks. As a result, in all these models from this time scale on the net feedback function $\widetilde{f}(p)$ is increasingly negative (not shown) and by eq. (4.8) the airborne fraction always decreases. In the HadGEM2-ES, at long time scales the magnitude of the land biogeochemical feedback starts to decrease, thereby decreasing the magnitude of the (negative) net feedback function and as a consequence increasing the airborne fraction.

By comparing subfigures (a) and (b) one can gain insight into the differences observed in the airborne fraction for the different models. For instance, the CESM1-BGC has the largest airborne fraction for all time scales. As seen in subfigure (b), this is probably to a large extent because at all time scales this model has clearly the smallest land biogeochemical feedback. As seen in last subsection, this behaviour is expected because this model accounts for nitrogen limitation, which reduces the CO_2 fertilization effect and thereby also the magnitude of the land biogeochemical feedback. As another example, on the lower end of the airborne fraction range are, among others, the MPI-ESM-LR and the BCC-CSM1-1. Looking at subfigure (b), one sees that the small values of their airborne fraction are to some extent explained by their large land biogeochemical feedback. At short time scales, this feedback is large but similar for both models. Accordingly, in this range the values of

6 This however does not work the other way around: From the positivity or negativity of the feedback functions one cannot determine the sign of the associated response functions. In fact, the only exception to the observation on the sign of the generalized sensitivities in Fig. 4.7 – the $\chi_\beta^{(L)}(t)$ sensitivity in the HadGEM2-ES – changes sign after 70 years but still leads to a negative $\widetilde{f}_\beta^{(L)}(p)$. But the negativity of $\chi_\beta^{(L)}(t)$ in the HadGEM2-ES is still reflected in Fig. 4.9b: Although for the domain analyzed $\widetilde{f}_\beta^{(L)}(p)$ is always negative, its magnitude starts to reduce over decadal time scales, which is in contrast to all other feedback functions.

their airborne fraction are small but lie close together. At time scales closer to 100 years, the MPI-ESM-LR has clearly the largest land biogeochemical feedback. Accordingly, at these time scales the MPI-ESM-LR presents the smallest airborne fraction, while the BCC-CSM-1 has an airborne fraction that is perceptively larger.

In the mean, the magnitude of the land biogeochemical feedback is always larger than that of the ocean biogeochemical feedback: At a time scale of 10 years, it is 1.4 times larger, and at a time scale of 100 years, 1.8 times larger. The picture is qualitatively similar for the radiative feedback: At a time scale of 10 years, the land feedback is 0.03 against approximately zero for the ocean feedback, and at a time scale of 100 years the land feedback is 7.4 times larger than its ocean counterpart. Aggregating land and ocean, the mean of the biogeochemical feedback is 22 times larger than the radiative feedback at a time scale of 10 years, and 5.6 times larger at a time scale of 100 years. These results suggest that at least on shorter time scales the difference between biogeochemical and radiative feedbacks may be even larger than suggested by previous studies employing Friedlingstein's framework (Gregory et al., 2009, Arora et al., 2013). However, one must note that here only the linear regime is considered so that the saturation of the land and ocean carbon sinks (that reduce the values of $\beta^{(L)}$ and $\beta^{(O)}$) is not taken into account.

Clearly, the model spread for the land feedbacks is much larger than that for the ocean feedbacks, as expected from previous studies (Gregory et al., 2009, Arora et al., 2013, Friedlingstein et al., 2014). The magnitude of this spread is better quantified in Fig. 4.10a, which shows at different time scales the standard deviation across all models for each feedback. As seen, the land biogeochemical feedback dominates the spread, with a value 4.5 times larger than the sum of the standard deviations for the other feedbacks at a time scale of 10 years, and 2.2 times larger than the sum at a time scale of 100 years. Because of the nonlinear relationship between $\widetilde{A}(p)$ and the feedback functions (see eq. (4.8)), how the model spread in the feedbacks propagates to the airborne fraction is not immediately clear. This propagation may be computed as (Roe, 2009)

$$\delta\widetilde{A}(p) = \widetilde{A}^2(p)\left(\delta\widetilde{f}_{\beta}^{(L)}(p) + \delta\widetilde{f}_{\gamma\alpha}^{(L)}(p) + \delta\widetilde{f}_{\beta}^{(O)}(p) + \delta\widetilde{f}_{\gamma\alpha}^{(O)}(p)\right) + h.o.t, \tag{4.26}$$

where $h.o.t.$ are higher-order terms. Fig. 4.10b shows the linear terms at the right-hand side of (4.26), the linear approximation of $\delta\widetilde{A}(p)$ (sum of the linear terms), and the true spread in the airborne fraction, taking for all variations one standard deviation. As seen, the true standard deviation in the airborne fraction follows closely the component arising from the land biogeochemical feedback, with a slightly larger discrepancy at time scales above 40 years. This indicates that most of the model spread in the airborne fraction arises from the spread in the land biogeochemical feedback. This result agrees with that obtained in a recent study (Jones and Friedlingstein, 2020) that also evaluated how the model spread in the airborne fraction is affected by the spread in the climate-carbon cycle feedbacks, but employing Friedlingstein's framework for the analysis.

A clearer understanding of the effect of the different feedbacks on the airborne fraction may be gained by changing the values of these feedbacks to simulate hypothetical situations

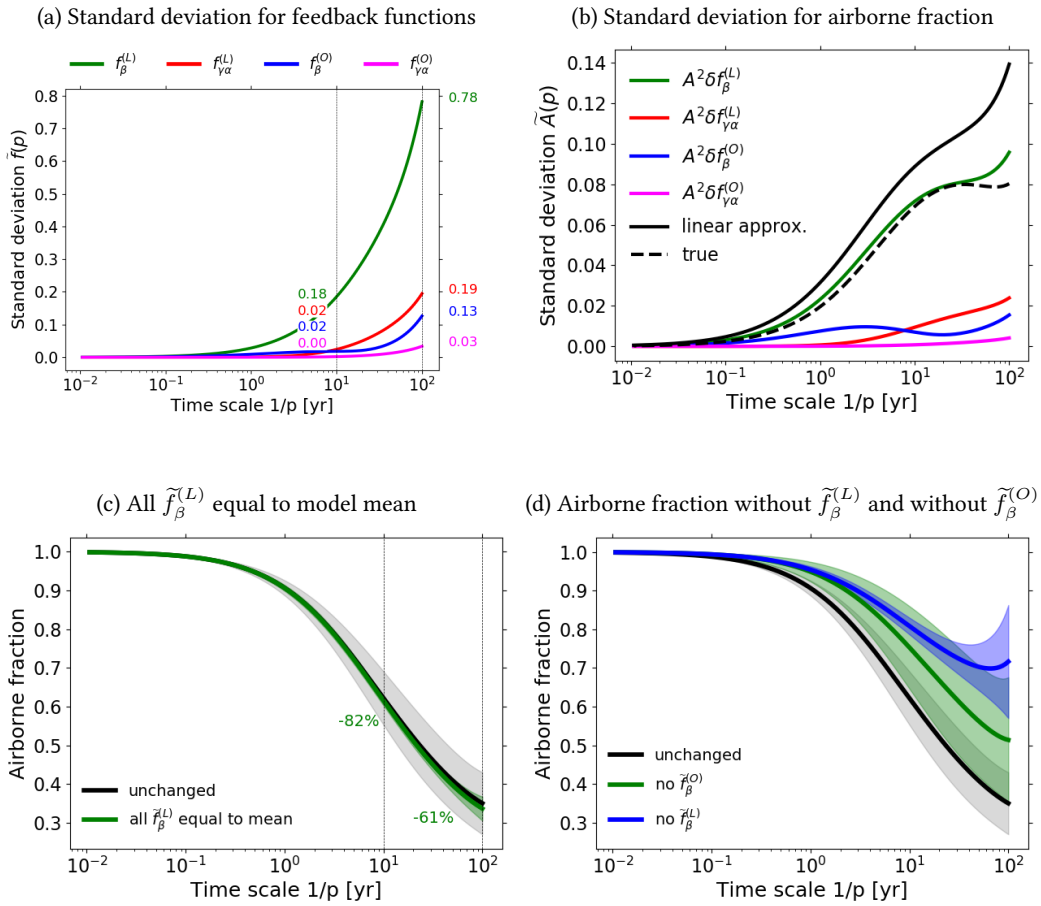


Figure 4.10 : Analysis of model spread and magnitude of feedback functions and their influence on the airborne fraction. (a) Standard deviation for each feedback function; (b) standard deviation for the airborne fraction and decomposition in terms of the feedback contributions according to eq. (4.26); (c) airborne fraction as presented in Fig. 4.9 (unchanged) and recalculated with the feedback function $\tilde{f}_\beta^{(L)}$ for all models set to the model mean; (d) airborne fraction as presented in Fig. 4.9 (unchanged) and recalculated with the land biogeochemical feedback for all models set to zero, and with the ocean biogeochemical feedback for all models set to zero. See text for more details.

and then evaluating the resulting change in the airborne fraction. For instance, one can illustrate how strongly the spread in the airborne fraction depends on the spread in the land biogeochemical feedback by recalculating the statistics of the airborne fraction taking $\widetilde{f}_{\beta}^{(L)}$ in all models as equal to its model mean. As shown in Fig. 4.10c, it turns out that if the exact values of only this feedback function were known and equal to the model mean (with all other feedback spreads remaining the same), the spread in the airborne fraction would reduce by about 82% at a time scale of 10 years and 61% at a time scale of 100 years.

With a similar procedure, one can separately compare the influence of the land and ocean biogeochemical feedbacks on the airborne fraction. Fig. 4.10d compares the statistics of the airborne fraction obtained in Fig. 4.9a with those of the airborne fraction calculated by excluding in all models the ocean biogeochemical feedback, and those of the airborne fraction calculated by excluding in all models the land biogeochemical feedback. As seen, by excluding the ocean biogeochemical feedback one obtains an airborne fraction with larger model spread (due to the large spread in the land biogeochemical feedback), larger mean values over all time scales (because of the absent contribution of the ocean biogeochemical feedback), but with a general behaviour close to the unchanged airborne fraction. On the other hand, by excluding the land biogeochemical feedback, the spread reduces for a large range of time scales (since the spread in the ocean biogeochemical feedback is much smaller), but the mean already on decadal time scales gets larger than in the previous case, changing even its qualitative behaviour at time scales close to 100 years: Instead of following the decreasing trend, the mean of $\widetilde{A}(p)$ starts to increase. This is because the ocean biogeochemical feedback shows relatively small contributions at these longer time scales (see Fig. 4.9b). Therefore, at these time scales contributions from the radiative feedbacks become relatively more significant, thereby changing the qualitative behaviour of the airborne fraction. This result emphasizes the large effect of the land biogeochemical feedback not only on the spread but also on the magnitude of the airborne fraction in these models.

4.6 SUMMARY

The dynamics of the global carbon cycle can be understood in terms of feedbacks arising in the land and ocean carbon cycle as atmospheric CO_2 is perturbed. To disentangle and separately quantify those feedbacks, Friedlingstein et al., 2003 developed the α-β-γ framework. Although this framework gives insight into the main effects of atmospheric CO_2 perturbations onto the global carbon cycle, it quantifies the feedbacks only for a particular perturbation scenario, and does not provide information on the internal dynamics of the system. Recent studies have proposed a generalization of this framework that characterizes the feedbacks independently of the perturbation scenario (Rubino et al., 2016, Enting and Clisby, 2019). The generalized framework assumes weak atmospheric CO_2 perturbations and characterizes the feedbacks at different time scales by means of linear response functions. As a result of the generalization, this framework not only quantifies the feedbacks for

any perturbation scenario but also gives a description of the linear dynamics of the global carbon cycle. Although this framework has been suggested, its applicability to Earth System Models had not been investigated so far.

Here, I studied its applicability to an ensemble of CMIP5 models. The purpose of the present study was twofold: 1) to demonstrate that the generalized α-β-γ framework consistently describes the linear dynamics of the global carbon cycle in an Earth System Model (the MPI-ESM); and 2) to quantify by this framework the climate-carbon cycle feedbacks and airborne fraction in different CMIP5 models. To achieve the first purpose and as a preparation for the second, the appropriate linear response functions (generalized sensitivities) were recovered for the MPI-ESM from standard 1% experiments employing the RFI method developed in chapter 2. Since the land carbon generalized sensitivities had been already recovered in chapter 3, here the generalized sensitivites for ocean carbon and temperature were recovered. The range of perturbation strengths that characterizes the linear regime for each response in the framework was estimated. As for the land carbon in chapter 3, the radiative response of ocean carbon was found to be linear for the whole 1% rad experiment. The responses of land and ocean temperature were estimated to be approximately linear within a range of CO_2 perturbation strengths of about 279 ppm. The biogeochemical response of ocean carbon was found to be linear for perturbations up to about 94 ppm, which is also approximately the linear regime for the biogeochemical response of land carbon. Because the biogeochemical responses presented the most restricted linear regime, it is their linear regime that limits the applicability of the generalized framework as a whole, which is thus valid for perturbations of up to about 94 ppm in the MPI-ESM. After recovering the generalized sensitivities, it was demonstrated that within the estimated linear regime they can be employed to predict the response of the MPI-ESM to different temporal developments of atmospheric CO_2 in concentration-driven simulations. In addition, to improve the quality of the recovered sensitivities, different approaches that account for known nonlinearities in the responses were considered. By employing these approaches, it was shown that the predictive power of the recovered generalized sensitivities can be extended: By taking as perturbation the logarithm of CO_2 instead of CO_2, the ocean biogeochemical response as well as land and ocean temperature responses were found to be approximately linear for the whole 1% experiment.

To demonstrate that the generalized framework correctly describes the dynamics of the carbon cycle in the MPI-ESM not only in concentration-driven simulations but also as a coupled system, results obtained from the framework were compared to the behaviour exhibited by the model in emission-driven simulations. Since it had been already demonstrated that the framework predicts the separate responses of land and ocean carbon to changes atmospheric CO_2, to demonstrate that it gives a correct description of the dynamics of the coupled system it was sufficient to show that it correctly predicts the response of atmospheric CO_2 to CO_2 emissions accounting for the climate-carbon cycle feedbacks. In the generalized framework, this response is determined from a generalized form of the airborne fraction that is time-scale dependent. Therefore, this airborne fraction was computed from the framework and compared to that obtained from emission-driven simulations. A

close agreement between the two estimates was found for the whole range of time scales analyzed, with an almost perfect agreement for time scales between 7 and 100 years. In addition, it was shown that both estimates agree with theoretical expectations.

Using the numerical approaches that gave the best results in the application to the MPI-ESM, the climate-carbon cycle feedbacks and airborne fraction were quantified for the ensemble of CMIP5 models. To this end, the RFI method was employed to derive the generalized sensitivities for all CMIP5 models using data from 1% experiments. The consistency of the obtained sensitivities was qualitatively confirmed by comparing the results to those obtained by Arora et al., 2013. Although the accuracy of the description of the carbon cycle dynamics by the generalized framework could not be thoroughly tested for all CMIP5 models due to the lack of appropriate additional simulations, the additivity of the biogeochemical and radiative responses that is assumed in the framework was shown to be a reasonable approximation within the estimated linear regime. With the generalized sensitivities, the time-scale dependent airborne fraction and climate-carbon cycle feedbacks were derived for each model. It was shown that in all models the airborne fraction is 1 for small time scales and for almost all models it strictly decreases as the time scale increases. At a time scale of 10 years the values range from 0.56 to 0.75 and at a time scale of 100 years they range from 0.26 to 0.5. It was also investigated how the dynamics of the airborne fraction is determined by the interplay of the climate-carbon cycle feedbacks. At small time scales the feedbacks approach zero; as a result, the airborne fraction approaches 1. At larger time scales, the radiative feedbacks are always positive and the biogeochemical feedbacks are always negative, with both generally increasing in magnitude as the time scale increases. Interestingly, for almost all models as the time scale increases the biogeochemical feedbacks are increasingly larger than the radiative feedbacks. As a result, in these models for increasing time scales the net climate-carbon cycle feedback gets increasingly negative and thereby the airborne fraction decreases.

Considering global carbon, in the model mean the biogeochemical feedback was found to be 22 times larger than the radiative feedback at a 10-years time scale and 5.6 times larger at a 100-years time scale. This result suggests that at least over shorter time scales the difference between these feedbacks may be even greater than previously thought (Gregory et al., 2009, Arora et al., 2013). Nevertheless, one must note that here only the linear regime is considered, so that the saturation of the land and ocean carbon sinks (that reduce the values of $\beta^{(L)}$ and $\beta^{(O)}$) is not taken into account.

The land feedbacks have larger model mean and spread than its ocean counterparts at all time scales. Particularly large is the spread in the land biogeochemical feedback: At a time scale of 10 years its standard deviation across models is 4.5 times larger than the sum of the standard deviations for the other feedbacks, while at a time scale of 100 years this ratio decreases to 2.2.

The influence of the spread and magnitude of the different feedbacks on the airborne fraction was also investigated. It was found that the spread in the airborne fraction arises mostly from the spread in the land biogeochemical feedback, especially for time scales

below 40 years. In the hypothetical case where this feedback were known and equal to the current model mean, while other feedbacks remained the same for each model, the spread in the airborne fraction would decrease by 82% at a 10-years time scale and by 61% at a 100-years time scale. Concerning the influence of the magnitude of the feedbacks, in the mean the land biogeochemical feedback demonstrated to have a greater influence on the magnitude and qualitative behaviour of the airborne fraction than the ocean biogeochemical feedback.

4.7 DISCUSSION

It is worth noting that while the generalized framework studied here was shown to reasonably describe the linear dynamics of the global carbon cycle in the MPI-ESM, the results obtained for the other CMIP5 models depend on two basic assumptions. The first is that the generalized sensitivities in the CMIP5 models were recovered with sufficient quality by the numerical approaches that were appropriate to recover the sensitivities in the MPI-ESM. This involves the assumption that all models have a similar linear regime to that found for the MPI-ESM for the different responses taken to recover the sensitivities. This might not be the case – and is probably not for $\chi_{\beta}^{(L)}$ in the HadGEM2-ES (see discussion of Fig. 4.7d). The second assumption is that the generalized framework itself correctly describes the dynamics of the global carbon cycle in those models. This involves the assumption that the biogeochemical and radiative responses are additive – which was confirmed to a good degree of approximation within the linear regime –, but also that the whole carbon cycle dynamics can be described in terms of the responses to atmospheric CO_2 and temperature given by eqs. (4.2)-(4.5). Ideally, both assumptions should be carefully investigated for each model with the aid of additional simulations, similarly to what was done for the MPI-ESM.

With that said, the conclusion that the spread in the airborne fraction arises mostly from the spread in the land biogeochemical feedback corroborates the recent finding by Jones and Friedlingstein, 2020, who performed a similar analysis employing Friedlingstein's framework. This agreement adds confidence in the results obtained here, and suggests that research on climate-carbon cycle feedbacks should focus especially on understanding the land biogeochemical feedback, since it is the largest source of model spread in the airborne fraction both here and in their analysis (which included also other model ensembles). But additionally, the conclusion drawn here is valid not only for a particular perturbation scenario, but for all scenarios within the linear regime. Insofar this conclusion addresses, for perturbations within the linear regime, the question raised by Jones and Friedlingstein, 2020 about the generalizability of their feedback analysis to scenarios with different characteristics.

Because the framework investigated here gives a description of the dynamics of the global carbon cycle, it may be applied not only to quantify the climate-carbon cycle feedbacks, but also to study, from the point of view of feedback analysis, a range of different dynamical problems. One of these problems, tackled here, is the description of the airborne fraction in

a way that is independent of the emissions scenario. It is known that the constant value of the airborne fraction depends on the particular exponential characteristic of the emissions curve (Raupach, 2013). As a result, once emissions deviate from this curve – as it must be if emissions are significantly reduced –, the airborne fraction is predicted to deviate from its constant value (Raupach, 2013). In such scenarios, the typical definition of the airborne fraction may be less insightful. Instead, the airborne fraction estimated here describes the response of atmospheric CO_2 to any emissions scenario, and gives the characteristics of this response at different time scales. While it is true that this form of the airborne fraction was obtained assuming weak perturbations, in chapter 3 and section 4.3 it was seen that the underlying nonlinearities in the carbon cycle are relatively weak, since they arise mostly from logarithmic dependencies. This suggests that for perturbations outside (but not very far from) the linear regime this generalized form of the airborne fraction should not change drastically, so that the estimates obtained here may still give valuable insight into the dynamics of the carbon cycle even in that case. Estimates of this generalized airborne fraction had already been attempted using eq. (4.21) (Enting, 2007, Enting and Clisby, 2019), but until now it had not been possible to link it to the dynamics of the climate-carbon cycle feedbacks, which requires the application of the generalized α-β-γ framework.

In direct connection with the airborne fraction is the Transient Climate Response to Cumulative Emissions (TCRE), which quantifies the change in global mean temperature in response to cumulative emissions (Ciais et al., 2013). As recognized by Gregory et al., 2009 and Jones and Friedlingstein, 2020, the TCRE can be investigated using the climate-carbon cycle feedback framework. In a similar analysis to that applied to the airborne fraction, Jones and Friedlingstein, 2020 quantified the contribution of the spread in each sensitivity from Friedlingstein's framework to the spread in TCRE, finding in their multi-model ensemble that the spread in the $\beta^{(L)}$-sensitivity has the second largest contribution, smaller only than that in the α-sensitivity (see their Fig. 4). With the framework studied here it would be interesting to try to check this finding in a more general setting, without additional complications arising from scenario dependence. Further, the generalized framework could be used to study how the TCRE is determined from the different climate-carbon cycle feedbacks at different time scales. Such analysis could lead to a better understanding of the dynamics behind this metric.

Additional applications of this framework may involve studying the contribution of the different feedbacks in the context of committed changes, where one is interested in understanding the behaviour of the system once emissions cease (Wigley, 2005, Plattner et al., 2008, Mauritsen and Pincus, 2017, MacDougall et al., 2020). Also, since the generalized framework gives a consistent formalism for quantifying climate-carbon cycle feedbacks, it may be possible to apply it to study in a consistent unified framework climate-carbon cycle feedbacks and physical climate feedbacks (Gregory et al., 2009, Williams et al., 2019, Goodwin et al., 2019). As pointed out by Jones and Friedlingstein, 2020, in principle the feedback framework opens also the possibility of constraining quantities such as TCRE and the airborne fraction from possible constraints on climate-carbon cycle feedbacks.

As for the problem investigated here, some ideas could in principle give more insight into the obtained feedbacks and airborne fraction. One idea could be to try to further understand the differences encountered across models by decomposing the feedback functions into terms related to processes involved in the different carbon responses. For instance, differences encountered in the land biogeochemical feedback may be associated with different sensitivities of plant productivity to CO_2, but can also be related to differences in the time-scales structure of the land carbon models that are forced by plant productivity. In chapter 3 it was shown that the land carbon generalized sensitivity $\chi_\beta^{(L)}$ can be obtained as a product of the sensitivity of Net Primary Production (NPP) to CO_2 and the response function that characterizes the response of land carbon to NPP. By this formulation of $\chi_\beta^{(L)}$ one could decompose the land biogeochemical feedback into these two factors, analyze the model spread for each factor and try to identify which factor mostly contributes to the spread in this feedback.

Another idea would be to try to investigate for different models the spectra of internal time scales of the response, as done for the land carbon in the MPI-ESM in chapter 3. This could hopefully lead to a better understanding of the main time scales underlying the different carbon responses of models, and maybe even allow for linking these differences to processes by their characteristic time scales. This however would require a careful investigation of the best approaches to derive the sensitivities in these models, as well as experiments with sufficiently high signal-to-noise ratio so that enough information about the spectra of time scales can be recovered.

Finally, one could try to obtain the airborne fraction in the time domain by inverting the Laplace transform obtained from eq. (4.8). Thereby one could hopefully perform a similar type of analysis to the one presented in this study, but with the advantage that the results in the time domain may offer a more direct physical interpretation.

5

CONCLUSIONS

In the introductory chapter of this dissertation I have set myself the goal of answering three sets of questions. In the present chapter, I summarize my results by answering those questions. Then, I provide a critical assessment of these results and discuss their implications in a wider context.

5.1 IDENTIFICATION OF LINEAR RESPONSE FUNCTIONS FROM ARBITRARY PERTURBATION EXPERIMENTS

In chapter 2 I have developed a method to identify linear response functions taking data from arbitrarily perturbed experiments. The robustness of the results by this method was demonstrated in several applications to data taken from experiments performed with a toy model. Here is the answer that I found for the research question investigated in this chapter:

Q2: How to identify linear response functions from arbitrary perturbation experiments?

A2: Identifying linear response functions from arbitrary perturbation experiments is an ill-posed problem. As such, it must be solved by a special method that accounts for this ill-posedness. My answer to this question is the RFI method. This method identifies the linear response function by taking data only from a single realization of an arbitrary perturbation experiment and a control simulation. The identification problem is solved by applying Tikhonov-Phillips regularization with the regularization parameter determined from Morozov's discrepancy method. The estimate of the noise level in the data required by the discrepancy method is given by a novel approach that combines an estimation of the high-frequency components of the noise from a spectral analysis of the data with an estimation of the low-frequency components of the noise from the control simulation. If the response function is known to be monotonic, this noise level can be further adjusted, which may improve the quality of the recovery of the response function. The RFI method recovers not only the response function but also the spectrum of time scales of the response, which may provide useful information about the internal dynamics of the system.

5.2 LINEAR RESPONSE FUNCTIONS AS GENERALIZATION OF LAND CARBON SENSITIVITIES IN THE MPI-ESM

The main goal of chapter 3 was to demonstrate the applicability of the RFI method to identify linear response functions from C^{4}MIP-type experiments performed with an Earth System Model – the MPI-ESM. The investigations in this chapter were not only a critical test for the RFI method, but also a test of the range of linearity for which the generalization of the land carbon sensitivities is valid. This range was identified by means of an additional technique that complemented the RFI method. Further, in this chapter it was shown how by recovering linear response functions with high quality one can obtain useful information about the internal dynamics of the carbon cycle. The answers I found for the research questions that motivated the study in this chapter are the following:

Q3.1: Can one robustly identify from C^{4}MIP-type experiments the linear response functions that generalize the land carbon sensitivities in the MPI-ESM? What is the range of linearity for which the generalization is valid?

A3.1: Yes. The generalized β- and γ-sensitivity for the land carbon were identified taking data from standard 1% experiments performed with the MPI-ESM. The robustness of the identified sensitivities was demonstrated by their ability to recover the response of the model in several experiments not used for the identification. The linear regime for the generalized γ-sensitivity was found to extend at least until the end of the 1% experiment, corresponding to temperature perturbations of about 6 K. For the generalized β-sensitivity, the range of linearity was found to extend up to 30 years in the 1% experiment, which corresponds to CO_2 perturbations of about 100 ppm.

It was found by the application of special approaches that the nonlinearity encountered in the land carbon response to CO_2 (characterized by the generalized β-sensitivity) is to a large extent explained by the nonlinear relationship between Net Primary Production (NPP) and CO_2. If one considers NPP instead of CO_2 as perturbation, the response is approximately linear for the whole 1% experiment. From the response function that characterizes this response one can as well derive the generalization of the β-sensitivity. The advantage of this more sophisticated approach is that since the response to NPP is linear up to larger perturbation strenghts, the quality of the data is higher and therefore the linear response function can be derived with higher quality. This information is relevant for the application of the generalized α-β-γ framework since it gives an approach to derive the generalized β-sensitivity for land carbon with better quality.

Likewise, it was found that the quality of the generalized γ-sensitivity can be improved by taking data from the $2\times CO_2$ experiment instead of a 1% experiment. The probable reason is that step experiments provide information on a wider range of frequencies than 1% experiments. This information is therefore also relevant for the application of the α-β-γ framework since it gives an additional approach by which one can improve the recovery of the generalized sensitivities.

Q3.2: Is it possible to obtain information about the internal dynamics of the carbon cycle from these functions?

A3.2: Yes. By the RFI method one can obtain the spectrum of time scales of the response. When such spectrum is obtained with high quality, it gives useful information about the internal dynamics of the system. This was the case for the best recovery of the generalized β-sensitivity (recovered with the special approach described in **A3.1** considering NPP instead of CO_2 as perturbation). The recovered spectrum showed peak time scales at about 4 and 100 years. Evidence for the robustness of this result was obtained by two independent tests.

I illustrated how one can investigate the reasons behind the structure of this spectrum by analyzing the hypothesis whether the two different peaks could be related to different carbon responses in the tropics and extra-tropics. The result turned out to be negative.

Nevertheless, the investigation of the responses for tropics and extra-tropics demonstrated some other characteristics of the system. The first is that, in agreement with expectations, the land carbon in the tropics responds faster than in extra-tropics. The second is that differences in the response functions for tropics and extra-tropics (that can also be interpreted as the impulse response of tropical and extra-tropical land carbon) arise from time scales between 60 and 250 years, with time scales above 250 years contributing mainly to differences in the offset between the response functions.

5.3 CLIMATE-CARBON CYCLE FEEDBACKS FOR WEAK PERTURBATIONS IN CMIP5 MODELS

In chapter 4 I investigated the generalized α-β-γ framework in Earth System Models. The purpose of this chapter was twofold. First, to demonstrate that the description of the linear dynamics of the carbon cycle given by this framework indeed agrees with the behaviour of the carbon cycle observed in an Earth System Model. Second, to apply this framework to quantify the generalized form of the climate-carbon cycle feedbacks and airborne fraction in different Earth System Models. The answers that I found for the research questions in this chapter are:

Q4.1: Can the generalized α-β-γ framework describe with reasonable accuracy the linear dynamics of the global carbon cycle in an Earth System Model?

A4.1: Yes. The accuracy of this description was demonstrated for the MPI-ESM. Evidences for this conclusion come from both chapters 3 and 4. First, in chapter 3 it was demonstrated that the generalized β- and γ-sensitivity for the land carbon can predict the responses of the MPI-ESM in several experiments. Then, in chapter 4 the same conclusion was reached for the generalized β- and γ-sensitivity for the ocean carbon, and for the generalized α-sensitivity for both land and ocean temperatures.

Collectively these results demonstrate the ability of the generalized sensitivities to describe the response of the model when atmospheric CO_2 perturbations are prescribed, i.e. in concentration-driven simulations. In addition, in chapter 4 it was also demonstrated that the generalized framework describes with reasonable accuracy the dynamics of the coupled carbon cycle in the MPI-ESM. The evidence for this conclusion was obtained by comparing the dynamics of atmospheric CO_2 – determined by the time-scale dependent airborne fraction – as predicted by the generalized framework against that observed in emission-driven simulations performed with the MPI-ESM. The prediction by the framework follows closely at different time scales the result obtained from simulations. Additional confidence in this result comes from the fact that the computed airborne fraction agrees with theoretical expectations.

Other relevant results from this part of the study are the following. For the generalized γ-sensitivity for the ocean carbon, a linear response was found for the whole 1% experiment. For the generalized β-sensitivity for the ocean carbon, the response was linear up to about 100 ppm. For the generalized α-sensitivity for land and ocean temperature, the response was found to be linear up to about 280 ppm. Therefore, when taking into account all generalized sensitivities, the β-sensitivites for both land and ocean present the most restricted linear regime. It is therefore their linear regime that determine the linear regime of the generalized framework as a whole. Hence, the range of linearity for which the generalized framework is valid in the MPI-ESM consists of CO_2 perturbations smaller than about 100 ppm.

Further, through the application of special approaches it was found that one can extend the predictive power of the sensitivities that showed a limited linear regime. By considering a logarithmic function of CO_2 as perturbation, it was found that the predictive power of both the β-sensitivity for the ocean and the α-sensitivity for land and ocean can be extended to the whole 1% experiment with a reasonable degree of approximation. These results might be relevant in applications where one is interested in employing separately these generalized sensitivities to study the biogeochemical response of ocean carbon, as well as the temperature responses over land and ocean.

Q4.2: How are the feedbacks and airborne fraction determined at different time scales in different Earth System Models? What can we conclude from them?

A4.2: The airborne fraction in all models is 1 at small time scales, which is expected both from theory (initial value theorem) and from understanding of the behaviour of the system: Because this airborne fraction determines the response of atmospheric CO_2 to emissions, at small time scales the airborne fraction must be 1, since then atmospheric CO_2 is simply equal to emissions. At longer time scales the situation is different. All models demonstrated an airborne fraction smaller than 1. For almost all models this airborne fraction strictly decreases as time scales increase. At 10-years time scale the range of values is 0.56-0.75, and at a 100-years time scale it decreases to 0.26-0.5. This decreasing behaviour is also in agreement with expectations from

theory (final value theorem) and understanding of the behaviour of the system: At longer time scales the land and ocean start to take up carbon so that the effect of emissions is buffered.

In contrast to the airborne fraction, in all models the feedbacks are zero at small time scales. This is also in agreement with understanding of the system: At small time scales, the land and ocean carbon cycle have not yet had time to respond to emissions, so that the feedbacks must be zero; this absence of feedbacks is what causes the airborne fraction to be 1 at small time scales. At larger time scales, biogeochemical feedbacks get negative, and radiative feedbacks get positive. This is in agreement with the current understanding that globally the biogeochemical feedbacks are related to sinks, and the radiative feedbacks to sources. But in addition, it turns out that the feedbacks are either positive or negative at all larger time scales, which is a non-trivial result. At each of these time scales the sum of the biogeochemical feedbacks is larger than that of the radiative feedbacks, resulting in a negative net feedback.

The most remarkable result from this investigation arises when analyzing the effect of the model spread in the feedbacks on the spread in the airborne fraction. It turns out that the spread in the airborne fraction is largely explained by the spread in the land biogeochemical feedback. In fact, in the hypothetical case where all models agreed on the values for this feedback, being these values equal to the current model mean, the resulting spread in the airborne fraction would reduce by 82% at a 10-years time scale and by 61% at a 100-years time scale. This result therefore suggests that research on the carbon cycle should focus especially on improving the understanding of the land biogeochemical feedback.

5.4 OUTLOOK

At the very beginning of this study I stated that this dissertation presents an *attempt* to systematically characterize the response of the global carbon cycle in Earth System Models. After all these investigations, can we say that this attempt was successful? In addition, what are the implications of this study within the wider scientific context? These questions have been to some extent addressed in the preceding chapters. In the following subsections I give an overview of the answers.

5.4.1 *A critical assessment of the results*

To answer the question whether my attempt in this dissertation was successful, we must critically assess the results obtained here. The idea of this study was to characterize the response of the carbon cycle by linear response functions. The particular focus was on those functions that quantify the climate-carbon cycle feedbacks (chapters 3 and 4); in addition,

the response of land carbon to Net Primary Production was also investigated (chapter 3). The main challenge in characterizing this response was how to identify the appropriate response functions from data that were not particularly suitable for this purpose. For this identification, a specific method had to be developed: The RFI method (chapter 1). Hence, this method is at the heart of the present study, and the answer to this first question lies in the robustness of this method.

Results from the RFI method were investigated in several ways. In chapter 2, the method was applied in many different situations to recover linear response functions taking data from experiments performed with a toy model. In general, the method demonstrated to give reasonable results, especially when monotonicity of the response function could be accounted for. Nevertheless, I found that the quality of the results depends critically on two aspects of the data: Noise and nonlinearity. This means that for the RFI method to give good results, the data must be sufficiently "clean" and contain no strong nonlinearities. In addition, when monotonicity cannot be accounted for, the quality of the recovery depends also on the validity of the spectral similarity assumption. Currently, no formal argument has been worked out to demonstrate in which cases this assumption holds, although intuitively it seems to be a reasonable assumption (see discussions in chapter 2).

Although in chapter 2 many tests were performed, there the RFI method was evaluated in the idealized situation where the "true" response function was known, and the conditions of noise and nonlinearity were controlled. This was not the case in chapter 3: In this chapter, the method was applied to identify the response functions that generalize the β- and γ-sensitivity for the land carbon taking data from C^4MIP-type experiments performed with the MPI-ESM. Also in this case, the method demonstrated to give reasonable results. Evidence for this conclusion was obtained by employing the recovered response functions to predict the response of the MPI-ESM in several experiments. Additional tests also indicated the robustness of the spectrum obtained from the best recovery of the generalized β-sensitivity.

Chapter 4 provided additional evidences for the robustness of the method. In this chapter, the method was employed to recover the response functions that generalize the α-β-γ sensitivities for the ocean carbon, land, and ocean temperature. Once more it was demonstrated that the recovered response functions could predict the response of the model in several experiments. Additionally, the recovered response functions, combined in the generalized α-β-γ framework, could very closely predict the airborne fraction obtained from emission-driven simulations at different time scales.

All these evidences thus indicate that the RFI method gave robust results for the generalized sensitivities in the MPI-ESM, and therefore that the global carbon response for weak perturbations was in this model reasonably well characterized. Nevertheless, as discussed in chapter 3, even with all these tests it must be emphasized that all these response functions are only approximations, so that one cannot be completely sure whether they can predict the response of the model to really all perturbation scenarios. This issue is not fundamental, but only technical: In principle, the quality of recovery of the response function – and therefore its predictive power – can be improved by taking data with higher signal-to-noise ratio, for instance by employing ensembles of experiments.

As for the characterization of the carbon cycle response in the other CMIP5 models (chapter 4), the picture is less clear. The reason is that for those models no extra experiments were available for careful tests to be carried out. Therefore, the quality of the characterization in those models depends on the assumptions that the approaches that worked best for the MPI-ESM are also appropriate to those models, and that the generalized α-β-γ framework gives an appropriate description for their carbon cycle dynamics. Ideally, such assumptions should be checked with the aid of additional experiments.

Still, there are some indirect indications that the results obtained for these models are robust. The first is that the qualitative behaviour of the generalized sensitivities agrees with their response characteristics (section 4.5.1). The second is that, within the estimated linear regime, in all models the additivity of the biogeochemical and radiative responses holds approximately. The third is that the characteristics of the resulting feedbacks and airborne fraction are in agreement with theoretical expectations and with the understanding of the system. And the fourth is that the result obtained here that the spread in the airborne fraction arises mostly from the spread in the biogeochemical feedback agrees with a recent finding by Jones and Friedlingstein, 2020, who performed a similar analysis but employing Friedlingstein's framework.

5.4.2 *Implications*

The main motivation for this study was the question whether the climate-carbon cycle feedbacks can be systematically quantified in Earth System Models by the generalization of the α-β-γ framework. On the way to investigate this question, the RFI method was developed. The considerations above suggest that both the RFI method and the generalized framework are appropriate for their purposes – to derive linear response functions from data, and to systematically quantify the feedbacks. But what can be learned from this generalized framework? How does the RFI method help?

From the point of view of the carbon-cycle community, the most relevant conclusion from the present study of the generalized framework is that the model spread in the airborne fraction can to a large extent be attributed to the spread in the land biogeochemical feedback. Although this result was obtained only for CMIP5 models, Jones and Friedlingstein, 2020 have also reached this conclusion by studying the CMIP5 and other model ensembles employing Friedlingstein's framework. This conclusion is relevant because it gives a direction where efforts must be concentrated if one wants to improve the knowledge of the airborne fraction. But if the same conclusion was also obtained by Jones and Friedlingstein, 2020, one could argue that a generalization of the α-β-γ framework was not needed. Although apparently the same conclusion was reached, there is one important difference that must be stressed: The result obtained here is general – at least if we restrict ourselves to the linear regime. In fact, Jones and Friedlingstein, 2020 emphasize the need to employ their approach to other types of scenarios to investigate whether their feedback analysis still holds. With the generalized framework, such extra investigations are not needed: The derived feedbacks

and airborne fraction are disentangled from the perturbation. Therefore, the conclusion that the spread in the airborne fraction is caused by the spread in the land biogeochemical feedback is not dependent on a perturbation scenario.

Looking ahead, the generalized framework offers opportunities to better understand the dynamics of the carbon cycle in terms of the climate-carbon cycle feedbacks in different situations. In their paper, Jones and Friedlingstein, 2020 investigate employing Friedlingstein's framework the influence of the different feedbacks on the Transient Climate Response to Cumulative Emissions (TCRE). In the same way, one can in principle apply the generalized framework to study this metric. The main advantage would be the same as for studying the airborne fraction: The analysis is disentangled from the type of perturbation, so that it is valid for all scenarios. In addition, the generalized framework would give insight into the dynamics of the system: In chapter 4 it was seen how the evolution of the feedbacks in the time-scale domain determines the different values of the airborne fraction. A similar analysis could be done for the TCRE, which could shed some light on the dynamics behind this metric.

Other potential application of the generalized framework is in studying committed changes. In this type of study one is interested in understanding how the system behaves after emissions cease (Wigley, 2005, Plattner et al., 2008, Mauritsen and Pincus, 2017, MacDougall et al., 2020). Such information depends on the memory of the system, and can be investigated by employing linear response functions. With the generalized framework, this type of study could be performed from the point of view of the different climate-carbon cycle feedbacks, i.e. one could study the influence of these feedbacks in the behaviour of the system in a scenario where emissions cease.

Recent studies have proposed to investigate in a unified framework climate-carbon cycle feedbacks and physical feedbacks (Gregory et al., 2009, Williams et al., 2019, Goodwin et al., 2019). In principle, one could employ the generalized framework to study this type of problem with the carbon cycle feedbacks being quantified in a systematic way. In addition, Jones and Friedlingstein, 2020 propose to employ their framework to constrain quantities such as the airborne fraction and the TCRE by constraining the feedbacks. This could in principle also be done with the generalized framework, with the advantage that the result would not be scenario-dependent.

More generally, the RFI method developed here can be useful to derive the response function in many problems where the linear response framework is employed. Examples of such problems are:

- Climate sensitivity: Ragone et al., 2016 showed how the concepts of equilibrium climate sensitivity and transient climate response can be understood as only particular values encoded in a linear response function. With the RFI method one could study such generalized framework for the climate sensitivity by deriving the appropriate linear response functions for different Earth System Models. This could provide insight into the time-scales structure of the response of climate to CO_2 perturbations;

- Prediction of the response of complex models: Lembo et al., 2020 showed by employing an ensemble of experiments how the linear response framework can be applied to predict the response of the MPI-ESM in many different climate variables. With the RFI method one could in principle extend this type of procedure to many Earth System Models without the need for ensembles of experiments. This could be useful in different studies where one needs to compute different scenarios with an efficient method;
- Understanding emergent constraints: Recent studies demonstrated how to obtain emergent constraints using response functions derived from conceptual models (Cox et al., 2018, Williamson et al., 2019). With the RFI method one could try to perform similar studies employing instead response functions derived from Earth System Models;
- Comparing spectra of time scales of different models: Since the RFI method gives not only the response function but also the spectrum of time scales of the response, in principle one can study the spectrum of time scales of different models. How such type of investigation could be done was already illustrated in chapter 3. In addition, in principle one could even try to compare the spectra of time scales from models to those from observations (e.g. Forney and Rothman, 2012b);
- Age and transit time distributions of carbon: Such characteristics can be directly obtained from linear response functions (Thompson and Randerson, 1999), so that with the RFI method one can derive these response functions and thereby these characteristics for different models;
- Changes in carbon turnover time following land use change: The linear response framework provides a formalism to systematically characterize the response of the terrestrial biosphere and carbon turnover times. By such framework one could study how changes in land cover can influence carbon turnover times and the response of land carbon to perturbations;
- Airborne fraction at different time scales from observations: As shown in chapter 4, by considering the linear response framework one can derive a time-scale dependent form of the airborne fraction. In principle by adapting the noise estimation procedure in the RFI method one could try to obtain this airborne fraction as well from observations. Comparing this airborne fraction to those obtained from models could provide insight into whether models are correctly capturing the response of atmospheric CO_2 to emissions at different time scales.

Many other applications are possible. In climate science, popularity of the linear response framework has been increasing over the last years (e.g. Lucarini, 2009, Lucarini and Sarno, 2010, Lucarini et al., 2014, Ragone et al., 2016, Lucarini et al., 2017, Aengenheyster et al., 2018, Ghil and Lucarini, 2019, Lembo et al., 2020, Bódai et al., 2020), and this framework has found applications in various fields like geoengineering and neurophysiology (Gottwald, 2020). In all these examples, as long as one has at hand a control and an arbitrarily perturbed

experiment, in principle the RFI method can be employed to derive the response function, which is the fundamental step in the practical application of this framework.

While it is true that the generalized α-β-γ framework studied here, and more generally, the linear response framework, are limited to the linear regime of the system, in principle these frameworks can be extended to higher-order terms (Ruelle, 1998b, Lucarini, 2009, Roe, 2009). But even when restricting ourselves to the linear regime, the results obtained in this dissertation give an example of the potential of this type of framework to provide a more general insight into different problems in carbon-cycle research and other fields.

A

APPENDIX TO CHAPTER 2

A.1 THE MAX PLANCK INSTITUTE FOR METEOROLOGY EARTH SYSTEM MODEL

In this appendix I give a brief description of the model employed in this study: the Max Planck Institute for Meteorology Earth System Model (MPI-ESM; for more details see Giorgetta et al., 2013). The MPI-ESM consists of the coupled general circulation models ECHAM6 (Stevens et al., 2013) for atmosphere, at T63/1.9° horizontal resolution with 47 vertical levels, and MPIOM (Jungclaus et al., 2013) for ocean, with a nominal resolution of 1.5° with 40 vertical levels. In addition, the MPI-ESM includes the subsystems JSBACH (Reick et al., 2013, Schneck et al., 2013), a land and vegetation model, and the marine biogeochemistry model HAMOCC (Ilyina et al., 2013). JSBACH and HAMOCC describe respectively the land and ocean carbon cycle in the MPI-ESM. Thus, of particular interest for our study is the subsystem JSBACH. JSBACH simulates fluxes of energy, water, momentum and CO_2 between the land surface and atmosphere. To represent subgrid scale heterogeneity, each grid cell of the land surface is divided into tiles, being each tile associated with a vegetation type (or "plant functional type"). The photosynthesis scheme follows Farquhar et al., 1980 and Collatz et al., 1992. The land carbon structure is divided into three vegetation pools (living tissues, carbohydrate and starch storage, and wood), four aboveground and belowground pools for litter from woody and nonwoody parts and a pool for soil carbon. In addition, JSBACH includes a dynamic vegetation scheme that allows for simulating changes in vegetation cover driven by climate.

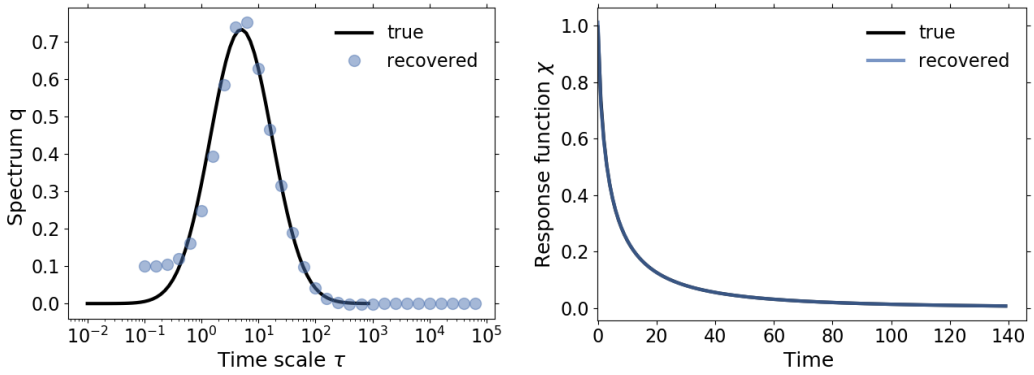
A.2 SENSITIVITY OF THE RECOVERED RESPONSE FUNCTION AND SPECTRUM TO THE PARAMETERS M, $\log \tau_{min}$ AND $\log \tau_{max}$ OF THE RFI ALGORITHM

In this appendix, it is shown that as long as the extent and resolution of the discrete distribution of time scales approximates the spectrum sufficiently densely, the derived spectrum $\mathbf{q}_\lambda$ and the derived linear response function $\chi(t)$ are approximately independent of the number of time-scales M and on the limits of the distribution $\log \tau_{min}$ and $\log \tau_{max}$. To isolate the effect of changes in M, $\log \tau_{min}$ and $\log \tau_{max}$ from the effect of noise, a relatively high $SNR \sim \mathcal{O}(10^5)$ is taken. For the computations I took data from 1% experiments with the toy model described in section 2.4.1. No monotonicity needed to be accounted for (step 6 of Fig. 2.2).

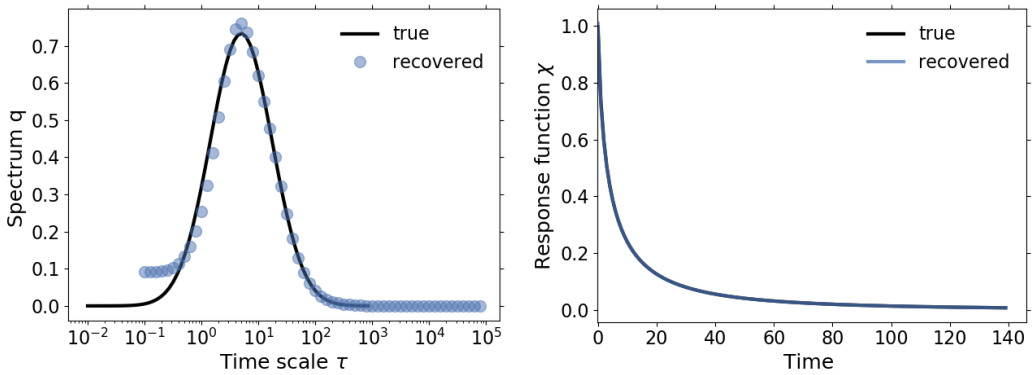
Figs. A.1 a-e show the recovery taking the same limits used throughout the paper ($\log \tau_{min}$ = -1 and $\log \tau_{max}$ = 5) but different number of time scales M. Figs. A.1 f-h show

the recovery keeping the number of time scales and the lower limit used throughout the paper (M = 30 and $\log \tau_{min}$ = -1) but changing the upper limit $\log \tau_{max}$. Figs. A.1 i-k show the recovery keeping the number of time scales and the upper limit used throughout the paper (M = 30 and $\log \tau_{max}$ = 5) but changing the lower limit $\log \tau_{max}$. As expected, the results are approximately independent of the changes in the prescribed parameters. The only substantial differences are found in the recovered spectra at time scales smaller than the time step $\Delta t = 1$, thus time scales over which anyway only little information is given by data. These small time scales are also problematic because of the ill-posedness of the problem that suppresses high-frequency information from the solution (see Groetsch, 1984, section 1.1).

(a) $M = 30, \log \tau_{min} = -1, \log \tau_{max} = 5$



(b) $M = 60, \log \tau_{min} = -1, \log \tau_{max} = 5$



(c) $M = 90, \log \tau_{min} = -1, \log \tau_{max} = 5$

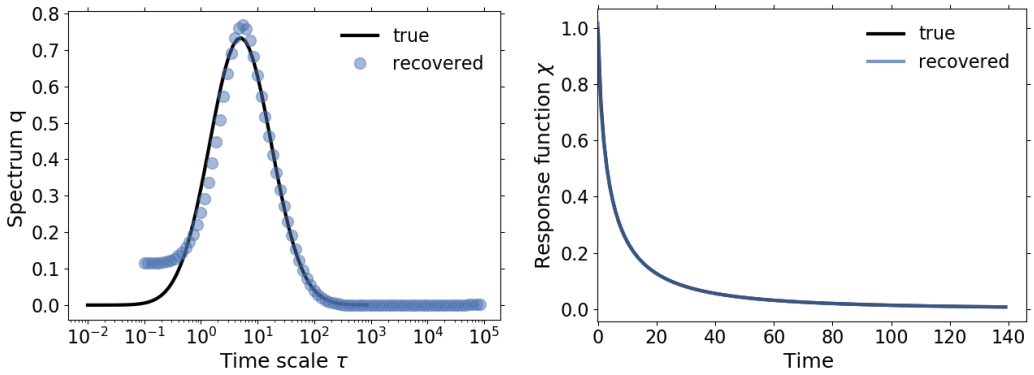
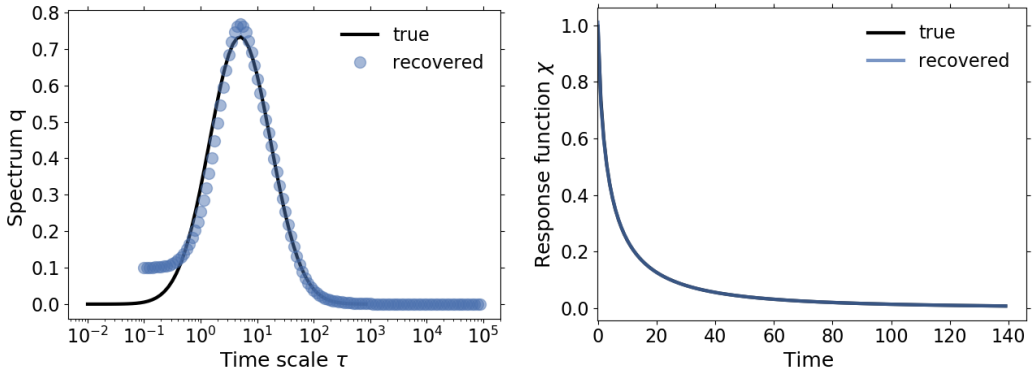
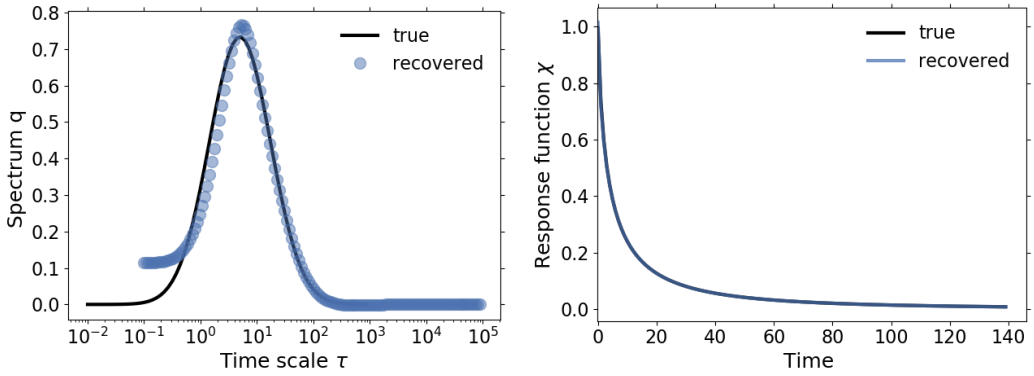


Figure A.1 : $\chi(t)$ and $\mathbf{q}_\lambda$ derived for different values of $M, \log \tau_{min}$ and $\log \tau_{max}$ ($SNR \sim \mathcal{O}(10^5)$). As long as the discrete distribution of time-scales approximates densely enough the spectrum, the derived $\chi_\lambda(t)$ and $\mathbf{q}_\lambda$ are approximately independent of the values of the prescribed parameters.

(d) $M = 120, \log \tau_{min} = -1, \log \tau_{max} = 5$



(e) $M = 140, \log \tau_{min} = -1, \log \tau_{max} = 3$



(f) $M = 30, \log \tau_{min} = -1, \log \tau_{max} = 7$

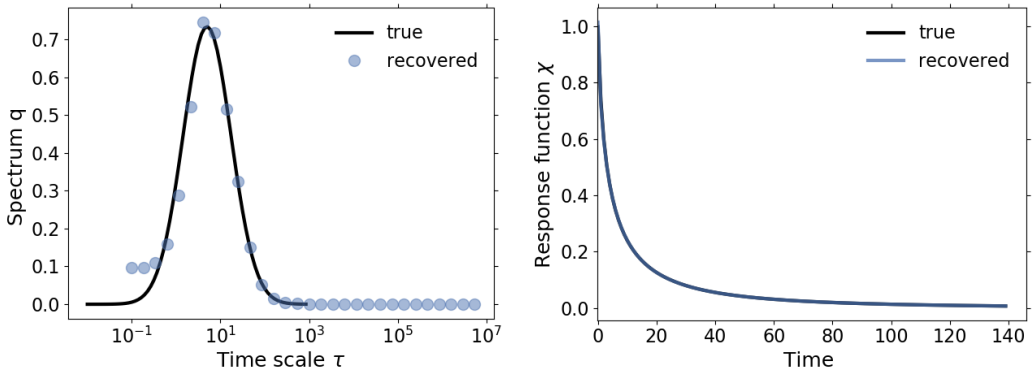
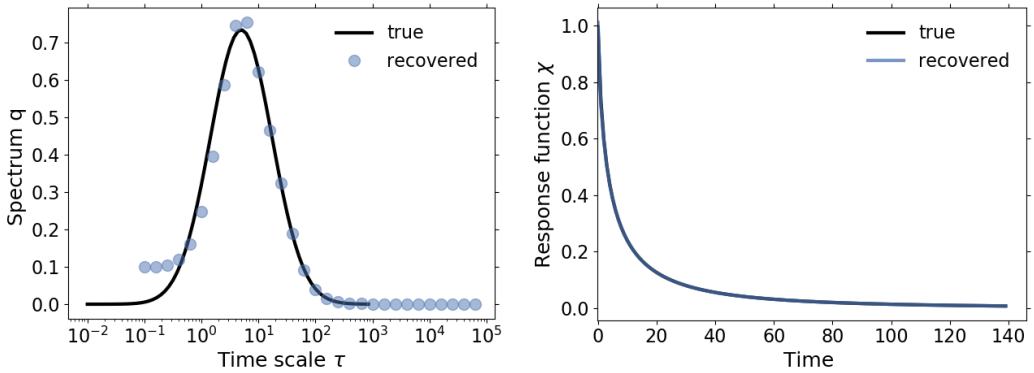
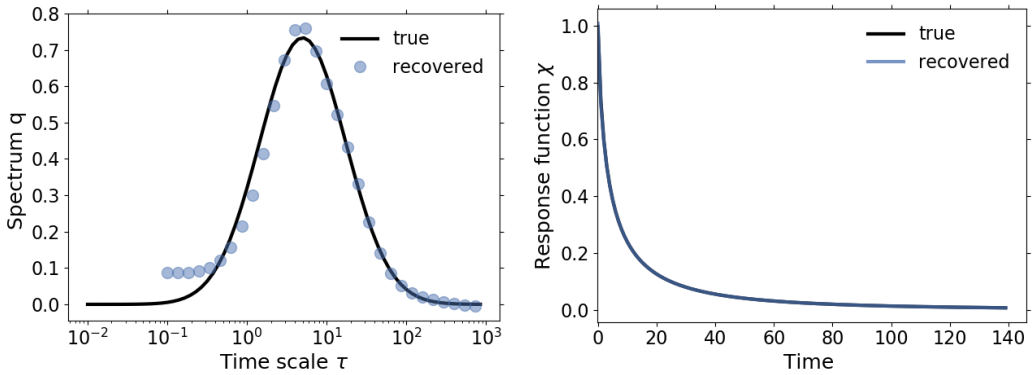


Figure A.1 : (continued) $\chi(t)$ and $\mathbf{q}_\lambda$ derived for different values of M, $\log \tau_{min}$ and $\log \tau_{max}$ ($SNR \sim \mathcal{O}(10^5)$). As long as the discrete distribution of time-scales approximates densely enough the spectrum, the derived $\chi_\lambda(t)$ and $\mathbf{q}_\lambda$ are approximately independent of the values of the prescribed parameters.

(g) $M = 30, \log \tau_{min} = -1, \log \tau_{max} = 5$



(h) $M = 30, \log \tau_{min} = -1, \log \tau_{max} = 3$



(i) $M = 30, \log \tau_{min} = -3, \log \tau_{max} = 5$

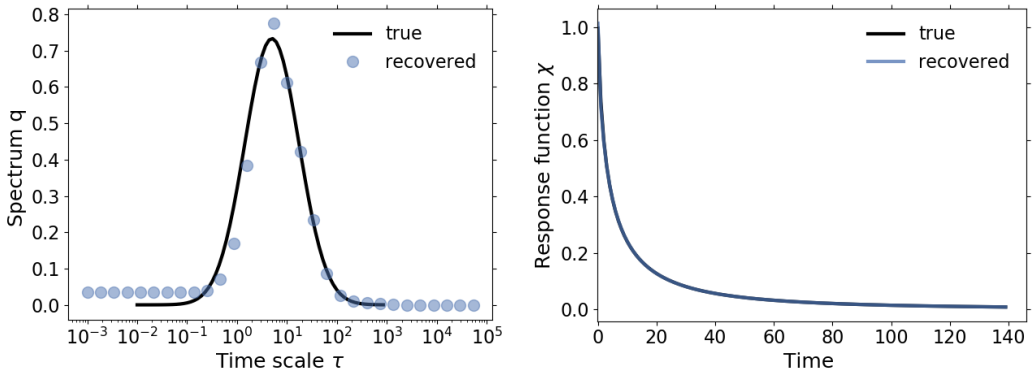
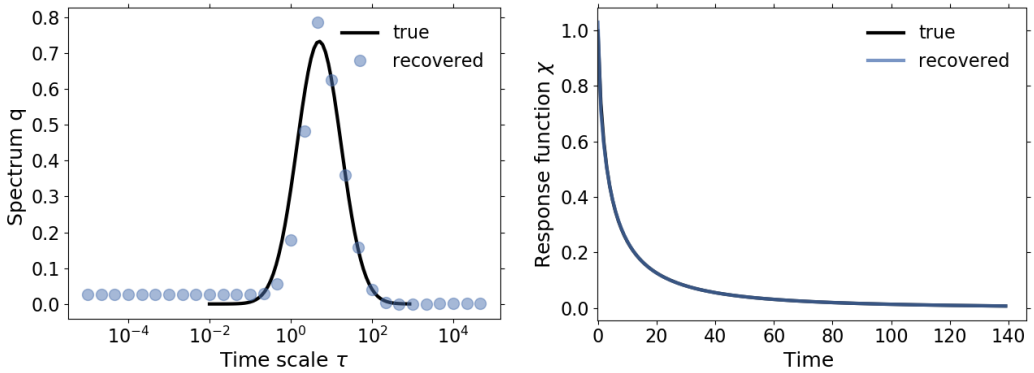


Figure A.1 : (continued) $\chi(t)$ and $\mathbf{q}_\lambda$ derived for different values of M, $\log \tau_{min}$ and $\log \tau_{max}$ ($SNR \sim \mathcal{O}(10^5)$). As long as the discrete distribution of time-scales approximates densely enough the spectrum, the derived $\chi_\lambda(t)$ and $\mathbf{q}_\lambda$ are approximately independent of the values of the prescribed parameters.

(j) $M = 30, \log \tau_{min} = -5, \log \tau_{max} = 5$



(k) $M = 30, \log \tau_{min} = -7, \log \tau_{max} = 5$

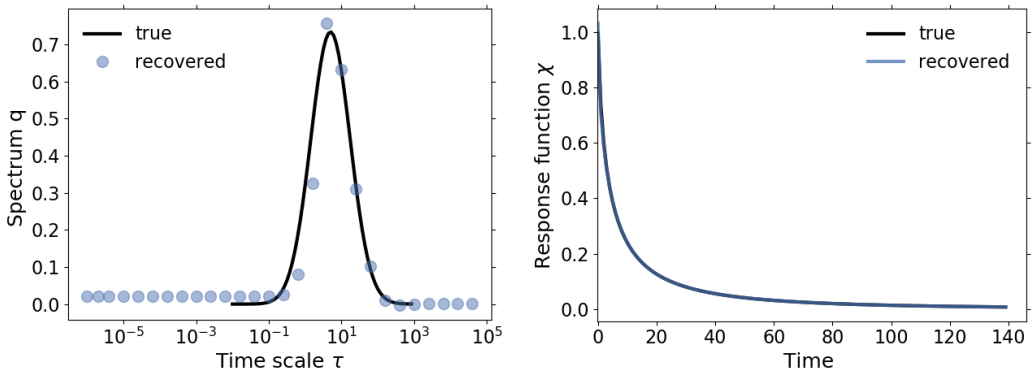


Figure A.1 : (continued) $\chi(t)$ and $\mathbf{q}_\lambda$ derived for different values of M, $\log \tau_{min}$ and $\log \tau_{max}$ ($SNR \sim \mathcal{O}(10^5)$). As long as the discrete distribution of time-scales approximates densely enough the spectrum, the derived $\chi_\lambda(t)$ and $\mathbf{q}_\lambda$ are approximately independent of the values of the prescribed parameters.

A.3 DERIVATION OF EQS. (2.14) AND (2.15) ON WHICH THE STUDY IS BASED

This appendix complements section 2.3.2 by deriving the set of eqs. (2.14), (2.15) underlying the RFI algorithm. They are a discretization of the basic definition (2.9) of the linear response function we are interested in. The special form (2.14), (2.15) involves in particular the logarithmic transformation (2.12) and a discretization of the representation (2.11) for the response function by means of a spectrum of time scales. Since $\chi(t)$ is assumed to be given by a spectrum of time scales according to (2.11), the discretization must be performed both in the time and time scale domain.

I start by defining the nondimensional time scale

$$\tau' := \frac{\tau}{\tau_0}, \tag{A.1}$$

where τ_0 is a reference time scale. Applying definition (A.1) in (2.11) gives

$$\chi(t) = \int_0^\infty g(\tau_0\tau')e^{-t/\tau_0\tau'}\tau_0 d\tau'. \tag{A.2}$$

Due to the wide range of time scales of the carbon cycle, calculations are facilitated if the time scales are evenly distributed at a logarithmic scale. To do so, the following change of variables is performed in (A.2):

$$\tau' = 10^z, \tag{A.3}$$

$$d\tau' = 10^z \ln 10\, dz = \tau' \ln 10\, d\log_{10}\tau'. \tag{A.4}$$

Thus, (A.2) becomes

$$\chi(t) = \int_{-\infty}^\infty g(\tau_0 10^{\log_{10}\tau'})e^{-t/\tau_0 10^{\log_{10}\tau'}}\tau_0\tau' \ln 10\, d\log_{10}\tau', \tag{A.5}$$

or simply

$$\chi(t) = \int_{-\infty}^\infty g(\tau_0\tau')e^{-t/\tau_0\tau'}\tau_0\tau' \ln 10\, d\log_{10}\tau'. \tag{A.6}$$

A convenient choice for the reference value is $\tau_0 = 1$ unit of time, so that by (A.1) the time scale $\tau = \tau'$ units of time. The resulting equation can thus be written as

$$\chi(t) = \int_{-\infty}^\infty q(\tau')e^{-t/\tau'} d\log_{10}\tau', \tag{A.7}$$

with

$$q(\tau') := \tau' \ln 10\, g(\tau'). \tag{A.8}$$

For convenience of notation we use simply τ instead of τ'.

For the discretization the support of $q(\tau)$ is assumed to lie within $[\log \tau_{min}, \log \tau_{max}]$. Accordingly, (A.7) reduces to

$$\chi(t) = \int_{\log \tau_{min}}^{\log \tau_{max}} q(\tau) e^{-t/\tau} d \log_{10} \tau. \tag{A.9}$$

Taking a constant step $\Delta \log \tau$ such that $\log \tau_{max} = \log \tau_{min} + M \Delta \log_{10} \tau$, (A.9) may be written as

$$\chi(t) = \sum_{j=0}^{M-1} \int_{\log \tau_{min} + j\Delta \log_{10} \tau}^{\log \tau_{min} + (j+1)\Delta \log_{10} \tau} q(\tau) e^{-t/\tau} d \log_{10} \tau. \tag{A.10}$$

Naming $t = (k+1)\Delta t$, eq. (2.9) can be rewritten as

$$\Delta Y(t) = \sum_{i=0}^{k} \int_{i\Delta t}^{(i+1)\Delta t} \chi(s) \Delta f(t-s) ds + \eta(t). \tag{A.11}$$

Plugging (A.10) into (A.11) and rearranging the resulting equation gives

$$\Delta Y(t) = \sum_{j=0}^{M-1} \int_{\log_{10} \tau_{min} + j\Delta \log_{10} \tau}^{\log_{10} \tau_{min} + (j+1)\Delta \log_{10} \tau} K(t, \tau) q(\tau) \, d \log_{10} \tau + \eta(t), \tag{A.12}$$

where

$$K(t, \tau) = \sum_{i=0}^{k} \int_{i\Delta t}^{(i+1)\Delta t} e^{-s/\tau} \Delta f(t-s) ds. \tag{A.13}$$

Assuming constant steps $\Delta \log_{10} \tau$ and Δt one may apply a quadrature rule (Hansen, 2002) to both (A.12) and (A.13), so that

$$\Delta Y(t) = \Delta \log_{10} \tau \sum_{j=0}^{M-1} K(t, \tau_j) q(\tau_j) + \varepsilon_\tau(t) + \eta(t), \tag{A.14}$$

$$K(t, \tau) = \Delta t \sum_{i=0}^{k} e^{-s_i/\tau} \Delta f(t - s_i) + \varepsilon_t(t, \tau), \tag{A.15}$$

where $\varepsilon_\tau(t)$ and $\varepsilon_t(t, \tau)$ are the errors resulting from the discretization. Plugging (A.15) into (A.14) yields

$$\Delta Y(t) \approx \Delta \log_{10} \tau \, \Delta t \sum_{j=0}^{M-1} \widetilde{q}(\tau_j) \sum_{i=0}^{k} e^{-s_i/\tau_j} \Delta f(t - s_i) + \eta(t) = \psi(t) + \eta(t), \tag{A.16}$$

where $\widetilde{q}$ is an approximation to q that accounts for the discretization errors. Now, if one requires that $\psi(t_k) + \eta(t_k) = \Delta Y(t_k)$ for particular times t_k,

$$\Delta Y(t_k) = \Delta \log_{10} \tau \, \Delta t \sum_{j=0}^{M-1} \widetilde{q}(\tau_j) \sum_{i=0}^{k} e^{-s_i/\tau_j} \Delta f(t_k - s_i) + \eta(t_k), \quad k = 0, 1, ..., N-1, \tag{A.17}$$

with the time steps chosen as follows

$$t_k = k\Delta t, \quad k = 0, 1, ..., N-1, \tag{A.18}$$

$$s_i = i\Delta t, \quad i = 0, 1, ..., k, \tag{A.19}$$

and the time scales

$$\tau_j = \tau_{min} 10^{j\Delta \log_{10}\tau}, \quad j = 0, 1, ..., M-1. \tag{A.20}$$

In order to simplify the notation, (A.17) is written as

$$\Delta Y_k = \Delta t \sum_{i=0}^{k} \chi_{k-i}\, \Delta f_i + \eta_k, \quad k = 0, ..., N-1, \tag{A.21}$$

$$\chi_k = \Delta \log_{10}\tau \sum_{j=0}^{M-1} q_j e^{-k\Delta t/\tau_j}, \quad k = 0, ..., N-1. \tag{A.22}$$

These are eqs. (2.14) and (2.15) underlying the study.

A.4 MONOTONICITY OF $\chi_\beta(t)$

In this appendix it is shown that the response function $\chi_\beta(t)$ is monotonic, as claimed in subsection 2.3.5 and chapter 3.1. The argument here is actually more general, namely that the response function to changes in the input of the system (for land carbon, the Net Primary Production) is monotonic. Since $\chi_\beta(t)$ is related to $\chi_{NPP}(t)$ by eq. (3.9), the monotonicity property transfers to $\chi_\beta(t)$.

Let the linear response of land carbon be described by eq. (3.5):

$$\Delta C^{bgc}(t) = \int_0^t \chi_{NPP}(t-s)\Delta NPP(s)ds, \tag{A.23}$$

where for simplicity we assume that $\eta(t)$ is small so that it can be neglected. If NPP is a Dirac delta function $\Delta NPP(t) = \delta(t)$, then the response is given by $\Delta C^{bgc}(t) = \chi_{NPP}(t)$. Therefore we can interpret the response function $\chi_{NPP}(t)$ as follows: If a certain number of carbon atoms enter the biosphere at time t = 0, the response function $\chi_{NPP}(t)$ gives the fraction of these atoms still left in the biosphere at time t.

Let $p(t)dt$ be the probability that an atom that entered the system at time t = 0 will leave it *at* time t. Then

$$P(t) := \int_0^t p(s)ds \tag{A.24}$$

is the probability that an atom that entered the system at time $t = 0$ will leave the system *until* time t. Hence from the interpretation of $\chi_{NPP}(t)$ above

$$P(t) = 1 - \chi_{NPP}(t). \tag{A.25}$$

From (A.24) and (A.25) it follows that

$$p(t) = -\frac{d}{dt}\chi_{NPP}(t). \tag{A.26}$$

But since the probability density function $p(t) \geq 0 \,\forall\, t$, then $\frac{d}{dt}\chi_{NPP}(t) \leq 0 \,\forall\, t$, i.e. $\chi_{NPP}(t)$ decays monotonically towards zero. Therefore, because $\chi_\beta(t)$ is simply a scaling of $\chi_{NPP}(t)$ given by eq. (3.9), $\chi_\beta(t)$ also decays monotonically towards zero.

A.5 SPECTRUM $q(\tau)$ POSITIVE OR NEGATIVE FOR ALL τ IMPLIES $\chi(t)$ MONOTONIC

This appendix is referred to on section 2.3.5 with the claim that a sufficient condition for $\chi(t)$ being monotonic is that all components q_i have the same sign. The proof is as follows.

Let $\chi(t)$ be defined by (2.12). Then,

$$\frac{d}{dt}\chi(t) = -\int_{-\infty}^{\infty} q(\tau)\frac{e^{-t/10^{\log_{10}\tau}}}{10^{\log_{10}\tau}} d\log_{10}\tau. \tag{A.27}$$

Since $10^{\log_{10}\tau} \geq 0$, $\frac{e^{-t/10^{\log_{10}\tau}}}{10^{\log_{10}\tau}} \geq 0 \;\forall\; t$. Thus, if $q(\tau) \geq 0 \;\forall\; \tau$, then $\frac{d}{dt}\chi(t) \leq 0 \;\forall\; t$. Similarly, if $q(\tau) \leq 0 \;\forall\; \tau$, then $\frac{d}{dt}\chi(t) \geq 0 \;\forall\; t$.

A.6 RESPONSE FUNCTION AND NOISE IN THE NONLINEARIZED RESPONSE FOR THE TOY MODEL

In this appendix it is shown how the linear response function and the noise terms are computed in section 2.4.5 when discussing by means of the toy model the complications arising from nonlinearity. I demonstrate that the linear response function for the nonlinear response ((2.44) with $a \neq 0$) of the toy model (section 2.4.1) can be analytically obtained from the linear case $a = 0$. Additionally, the noise from the control simulation and the combined noise in the response are defined.

I first demonstrate how to obtain the linear response function. Plugging (2.35) into (2.44) gives

$$Y_{nonlin}(t) = [1-2a\eta^*(t)]\int_0^t \chi^*(t-s)f(s)ds + \eta^*(t)[1-a\eta^*(t)] - a\left(\int_0^t \chi^*(t-s)f(s)ds\right)^2. \tag{A.28}$$

Taking the ensemble average of (A.28) and noting that $\langle \eta^*(t) \rangle = 0$ gives

$$\langle Y_{nonlin}(t) \rangle = \int_0^t \chi^*(t-s) f(s) ds + \mathcal{O}(f^2). \tag{A.29}$$

Therefore, $\chi^*(t)$ obtained for $a > 0$ from the nonlinearized response (2.44) is the same as for the case $a = 0$.

Now, by taking $f = 0$ in (A.28) one obtains for this nonlinear case the noise from the control simulation:

$$\eta_{ctrl}(t) := \eta^*(t)[1 - a\eta^*(t)]. \tag{A.30}$$

To define the combined noise $\eta(t) + \widetilde{\eta}(t)$, one must first define the nonlinear term $\widetilde{\eta}(t)$ from (2.42). For the nonlinearized response from the toy model, this term is given by the nonlinear term in (A.28), i.e.

$$\widetilde{\eta}(t) := -a \left(\int_0^t \chi^*(t-s) f(s) ds \right)^2. \tag{A.31}$$

Then, the noise term consists of the remaining terms of the nonlinear response Y_{nonlin} after subtracting the "clean" linear response and the nonlinear term $\widetilde{\eta}$, i.e.

$$\eta(t) := Y_{nonlin}(t) - \int_0^t \chi^*(t-s) f(s) ds - \widetilde{\eta}(t) = -2a\eta^*(t) \int_0^t \chi^*(t-s) f(s) ds + \eta^*(t)[1 - a\eta^*(t)]. \tag{A.32}$$

Hence, the combined noise is given by

$$\eta(t) + \widetilde{\eta}(t) := -2a\eta^*(t) \int_0^t \chi^*(t-s) f(s) ds + \eta^*(t)[1 - a\eta^*(t)] - a \left(\int_0^t \chi^*(t-s) f(s) ds \right)^2. \tag{A.33}$$

B

APPENDIX TO CHAPTER 3

B.1 DERIVATION OF SPECTRUM AND $\chi(t)$ WHEN THE RESPONSE CONTAINS TIME SCALES MUCH LONGER OR MUCH SHORTER THAN THE TIME SCALES COVERED BY DATA

Time scales much longer or much shorter than the time scales covered by data cannot be correctly recovered in the spectrum. Nevertheless, in this appendix I show that the wrong recovery of these extreme time scales does not strongly affect the recovery of $\chi(t)$. These considerations add to the footnote remark in subsection 3.5 where I claim that such extreme time scales can be safely ignored.

First, I consider the case where the response has time scales much longer than those covered by data. Let χ at time T be given by

$$\chi(T) = \int_{-\infty}^{\infty} q(\tau) e^{-T/\tau} d\log_{10}\tau. \tag{B.1}$$

Let T be the time series length and assume that $q(\tau)$ has significant contributions at time scales $\tau \geq \tau_L$ with $\tau_L >> T$. Then (B.1) can be written as

$$\chi(T) \approx \int_{-\infty}^{\log_{10}\tau_L} q(\tau) e^{-T/\tau} d\log_{10}\tau + \int_{\log_{10}\tau_L}^{\infty} q(\tau) d\log_{10}\tau, \tag{B.2}$$

where $e^{-T/\tau} \approx 1$ was used in the second integral because $\tau_L >> T$. Thereby the second term in the right-hand side of (B.2) is just a constant offset

$$\int_{\log_{10}\tau_L}^{\infty} q(\tau) d\log_{10}\tau = k, \quad k \text{ constant}. \tag{B.3}$$

Hence, for internal time scales much longer than the time series length T, the correct recovery of the individual $q(\tau)$ values is not relevant for the derivation of $\chi(t)$, as long as those values combine to the correct offset in eq. (B.3). Note that because this argument is based on the condition $\tau_L >> T$, it applies not only to $\chi(T)$ but also to all $\chi(t)$, $t < T$.

Now, I consider the case where the response has time scales much smaller than those covered by data. Let $t =: i\Delta t$, where Δt is the time step and $i \in IN$. If $q(\tau)$ has significant contributions at time scales $\tau \leq \tau_S$ with $\tau_S << \Delta t$, then for $i > 0$ eq. (2.12) can be written as

$$\chi(i\Delta t) \approx \int_{\log_{10}\tau_S}^{\infty} q(\tau) e^{-i\Delta t/\tau} d\log_{10}\tau, \tag{B.4}$$

where $e^{-i\Delta t/\tau} \approx 0$ was used for $\tau < \tau_S$ because $\tau_S << \Delta t$. As a result, values of $q(\tau)$ are irrelevant to $\chi(t)$ for almost every i. The only exception is $i = 0$, where one has

$$\chi(0) = \int_{-\infty}^{\infty} q(\tau) d \log_{10} \tau. \tag{B.5}$$

Hence, for time scales much shorter than the time step Δt, the correct recovery of the individual $q(\tau)$ values is not relevant for the derivation of $\chi(t)$ as long as those values combine with the other recovered values of $q(\tau)$ to the correct $\chi(0)$ in (B.5). As shown by (B.4), if they combine to the wrong value, only the recovery of $\chi(0)$ is affected.

Therefore, in principle $\chi(t)$ can be correctly recovered also when the response contains much longer and much shorter time scales than those covered by data, as long as the recovered $q(\tau)$ values at these extreme time scales combine to the correct values of the offset in eq. (B.3) and $\chi(0)$ in eq. (B.5). In Fig. B.1 I show an example of recovery of the spectrum and response function in such a case. For the recovery I took data with $SNR = 6 \times 10^4$ from the 1% experiment with the toy model (section 2.4.1). As seen, the spectrum can only be partially recovered (subfigure (a)). The recovery wrongly estimates spectral contributions at time scales longer and shorter than those covered by data, i.e. time scales larger than the time series length T = 140 and smaller than the time step Δt = 1. In addition, the existence of such spectral contributions at time scales not covered by data seems to also deteriorate the recovery of the spectrum at the time scales actually covered by data: In the region $1 < \tau < 140$, despite the high SNR the recovered spectrum matches the true spectrum only partially. This is probably a compensation effect where wrong information shows up in the recovered spectrum to compensate missing information on the response function. For instance, to obtain the correct $\chi(0)$, only (B.5) is needed but not the correct recovery of all time scales. The same goes for obtaining the correct offset: Only (B.3) is needed but not the correct recovery of all time scales from τ_L onwards.

Subfigure (b) shows the quality of the recovery of the response function $\chi(t)$. The recovered $q(\tau)$ values at long time scales, although individually wrong, combine to the correct offset in (B.3): The "height" of the recovered response function matches almost perfectly that of the true response function (compare e.g. the value at t = 140 for the true and recovered response function). On the other hand, $\chi(0)$ is not perfectly recovered (compare the difference at t = 0 between the true and recovered response functions), meaning that the sum in eq. (B.5) is incorrect. Nevertheless, the recovered value is still reasonably close to the true value. Except for this small error at $\chi(0)$, the overall recovery of the response function is almost perfect.

Therefore, as claimed in subsection 3.5, this numerical example shows that even though very long and very short time scales cannot be correctly recovered in the spectrum, they do not strongly influence the recovery of $\chi(t)$. This is because they only influence the offset in (B.3) and $\chi(0)$ in (B.5), and those two values seem to be reasonably well recovered numerically.

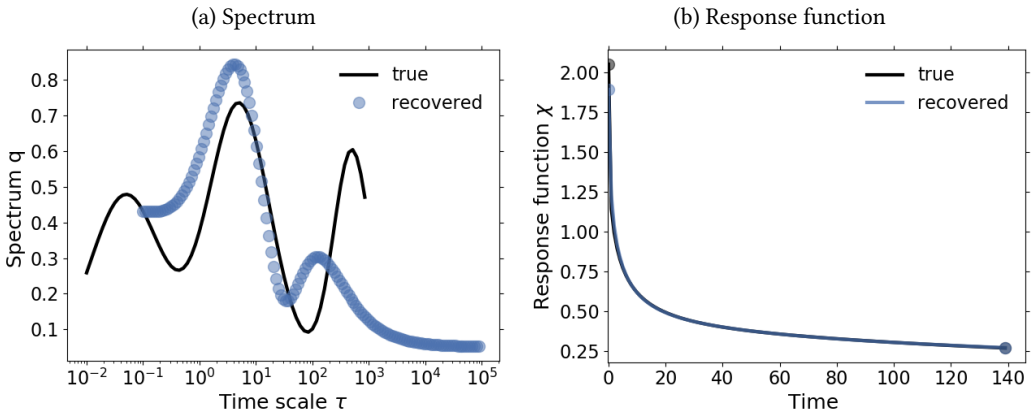


Figure B.1 : Recovery of spectrum and $\chi(t)$ taking a true spectrum with contributions at time scales much longer and much shorter than the time scales covered by data. For the recovery, data with $SNR = 6 \times 10^4$ from the 1% experiment with the toy model (section 2.4.1) was taken (times series length T = 140 and time step Δt = 1). Parameters are taken as in section 2.4.2 except for $M = 140$. The derived spectrum recovers only partially the true spectrum. Nevertheless, the response function $\chi(t)$ is almost perfectly recovered.

B.2 RECOVERING A DISCRETE SPECTRUM $q(\tau)$

This appendix gives numerical evidence for the claim in chapter 3.6 that although the RFI method assumes that the response has a continuous spectrum of many time scales, for sufficiently good signal-to-noise ratio the recovered continuous spectrum may also give a reasonable approximation to an underlying discrete spectrum of only few time scales.

Before I give numerical examples, one must first understand the limitations of the recovery of the spectrum. As explained in section 2.3.3, the spectrum $g(\tau)$ can only be completely recovered (for sufficiently high SNR) if it is dominated by the first components of expansion (2.23). By Hansen's observation (see section 2.3.3), this means that the spectrum must be dominated by low frequencies, i.e. it must be to some extent smooth. However, an underlying discrete spectrum implies a spectrum that is instead given by "spikes" in the time scale domain. Such "spikes" can only be described by high-frequency components of (2.23). Therefore, a discrete spectrum can be only partially recovered. To obtain a sufficiently good recovery, the solution (2.23) must contain many components. Hence, the data must have a sufficiently high signal-to-noise ratio. The explanation for this conclusion is the following: If many components in (2.23) must be recovered, the regularization parameter must be small. But a small regularization parameter requires a small noise level δ (see Theorem 3.3.1 in Groetsch, 1984 or explanation number 4. in section 2.3.5).

With this in mind, I show in Fig. B.2 that at least smooth approximations to an underlying discrete spectrum can be obtained. For the results I took data with $SNR \sim \mathcal{O}(10^4) - \mathcal{O}(10^5)$

from different experiments performed with the toy model described in section 2.4.1. Since the aim is to recover discrete spectra, a larger number of time scales M gives a better resolution. Therefore, I take $M = 140$. All other parameters are taken according to section 2.4.2. Also, monotonicity needed not to be taken into account (step 6 of Fig. 2.2).

Subfigures (a)-(f) show results for taking data from a 1% experiment and subfigures (g)-(j) from a $2\times f_0$ experiment. I start with one time scale $\tau = 37$ (subfigure a). As expected, the spike cannot be perfectly recovered, but the recovery gives a smooth approximation to it, with peak coinciding approximately with the "true" value. On the other hand, the response function is almost perfectly recovered. This is a result of the ill-posedness of eq. (2.12): In the same way that high frequencies of the spectrum are suppressed in the data by eq. (2.1), they are also suppressed in the response function by eq. (2.12) (see Groetsch, 1984, section 1.1). Therefore, both the true spectrum and its smooth recovery result in practically the same response function.

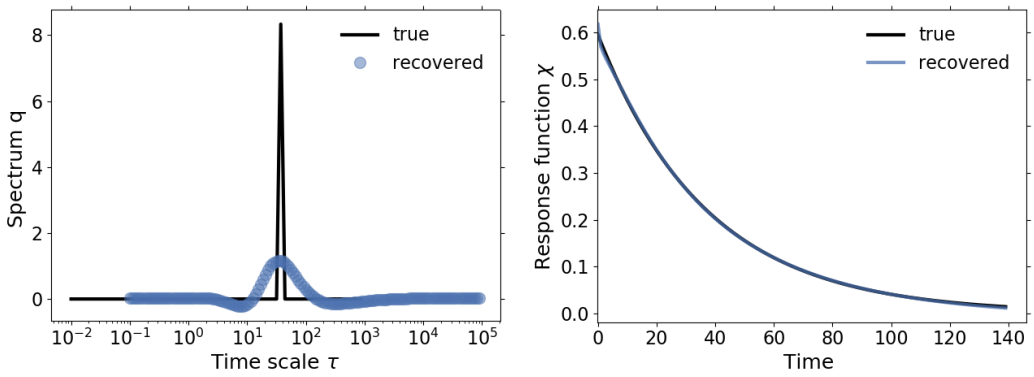
In subfigure (b), the true spectrum has two time scales, this time at $\tau = 7$ and $\tau = 100$. Similarly to subfigure (a), the time scales are recovered by a smooth approximation with peaks approximately at the true values. Also similarly to subfigure a, the response function is almost perfectly recovered. A similar result is obtained if we take time scales that are a bit closer together, as seen in subfigure (c) ($\tau = 7$ and $\tau = 37$). Nevertheless, now the peak for the longer time scale is a worse approximation, and there is a slightly pronounced negative peak that does not reflect the true spectrum.

In subfigure (d), I take instead time scales $\tau = 37$ and $\tau = 100$ (subfigure d). Here, the resolution is not sufficiently high for a recovery of each time scale separately. Instead, the recovered spectrum shows only one mode that spans both spikes. Once more the response function can be almost perfectly recovered. Taking the three time scales $\tau = 7$, $\tau = 37$ and $\tau = 100$ (subfigure e), the smooth recovery shows only two modes: One at $\tau = 7$ with peak almost coinciding with the true value, and another with peak in between $\tau = 37$ and $\tau = 100$. If one in addition considers a fourth time scale $\tau = 1$ (subfigure f), once more the recovery shows only two modes. But now the mode at shorter time scales displays a longer tail that spans both $\tau = 7$ and $\tau = 1$.

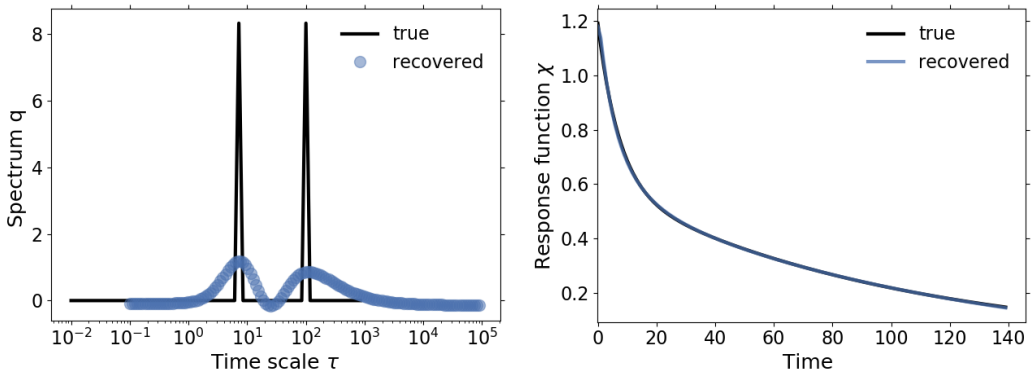
The situation changes when we take for the recovery the $2\times CO_2$ experiment. According to the Laplace transform analysis in section 3.3, data from this experiment should give more information on small time scales. Subfigure (g) shows that this is indeed the case: In contrast to the recovery from subfigure (f), now the time scale $\tau = 1$ is also recovered. On the other hand, now there are two small negative peaks that do not reflect information from the true spectrum. Also, the resolution is not sufficiently good to recover separately the two peaks at long time scales. However, if the longest time scale is increased from $\tau = 100$ to $\tau = 228$ (subfigure h), a small peak can be seen around this time scale. In addition, now the peak at $\tau = 37$ matches slightly better the true value. In subfigure (i) I add another time scale, now at $\tau = 518$. We see that although this time scale cannot be recovered separately from $\tau = 228$, the peak at long time scales is more pronounced. This is in contrast to the fact that the time series used for the recovery reaches only t = 140, indicating that the

method can recover information on time scales even longer than the time series length. This is in agreement with the conclusions from section 3.4, where in one case $\chi_\beta(t)$ was recovered from a time series with only 30 years but could recover responses for t = 140 years (see Fig. 3.7b).

(a) $\tau = 37$



(b) $\tau = 7, \tau = 100$



(c) $\tau = 7, \tau = 37$

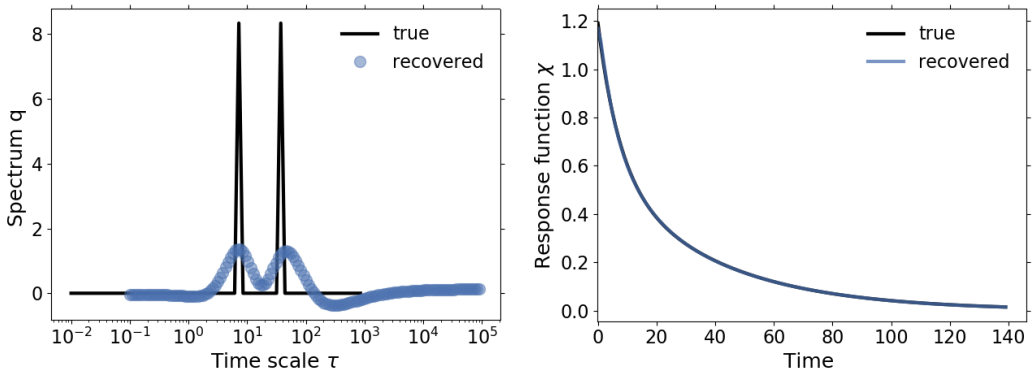


Figure B.2 : Spectrum $\mathbf{q}_\lambda$ and response function $\chi(t)$ derived for different discrete spectra $\mathbf{q}^*$ for $SNR \sim \mathcal{O}(10^4) - \mathcal{O}(10^5)$. Subfigures (a)-(f) show results for taking data from a 1% experiment and subfigures (g)-(j) for a $2\times f_0$ experiment. The recovered spectra give smooth approximations to the underlying discrete spectra. The response functions are recovered almost perfectly.

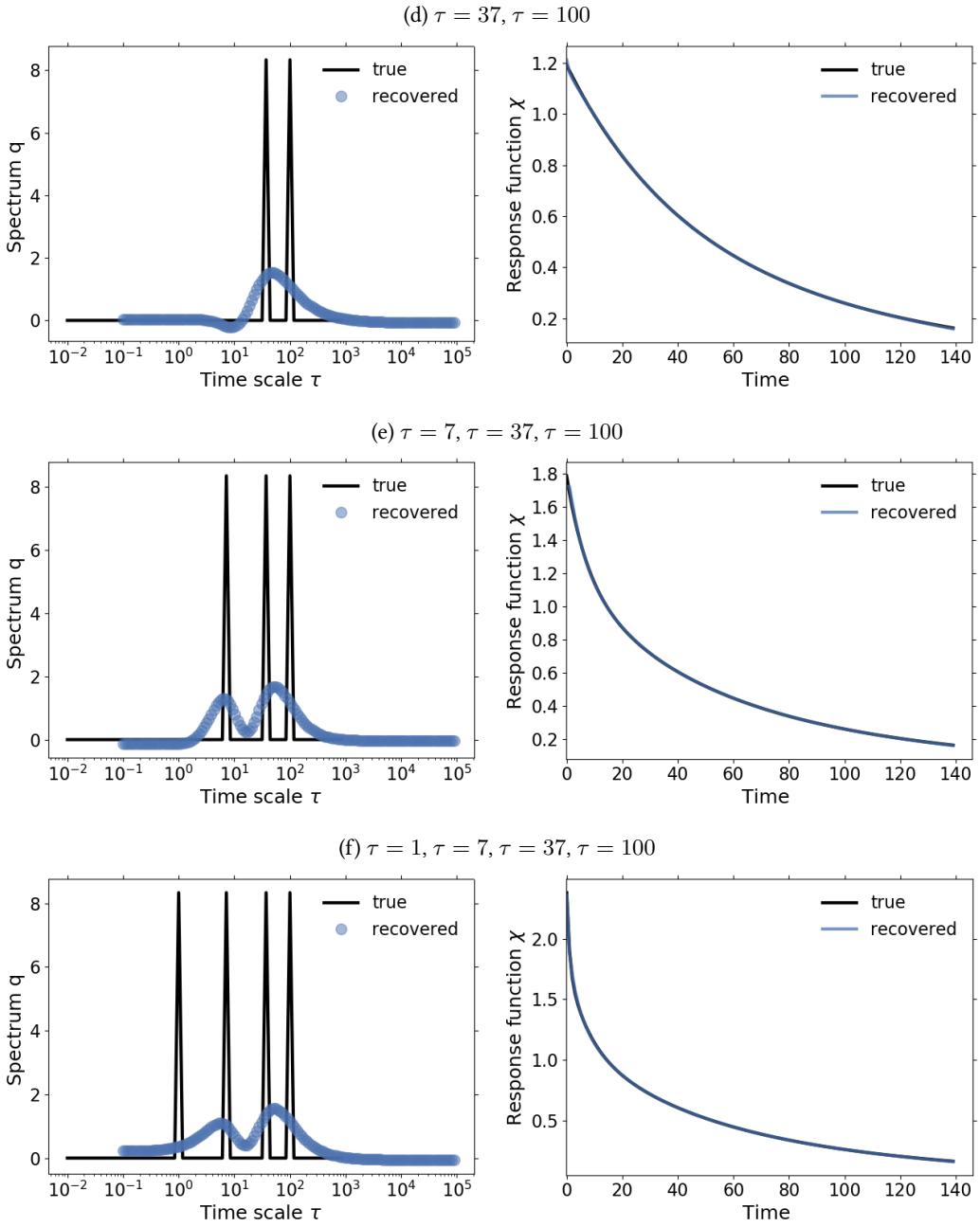
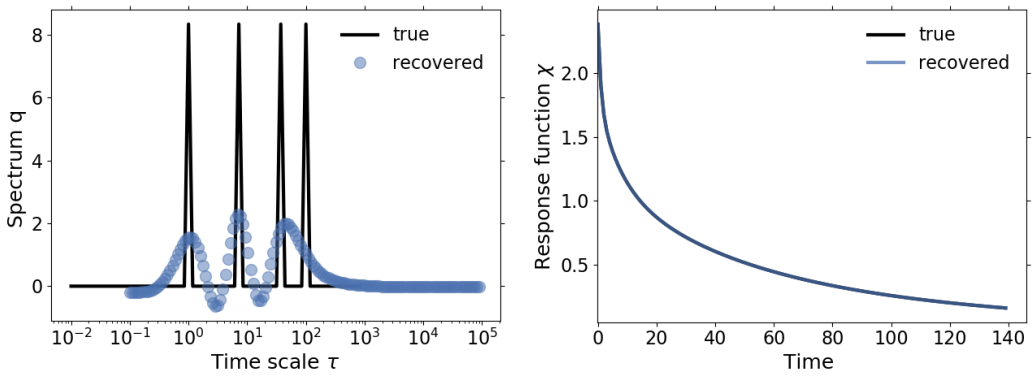
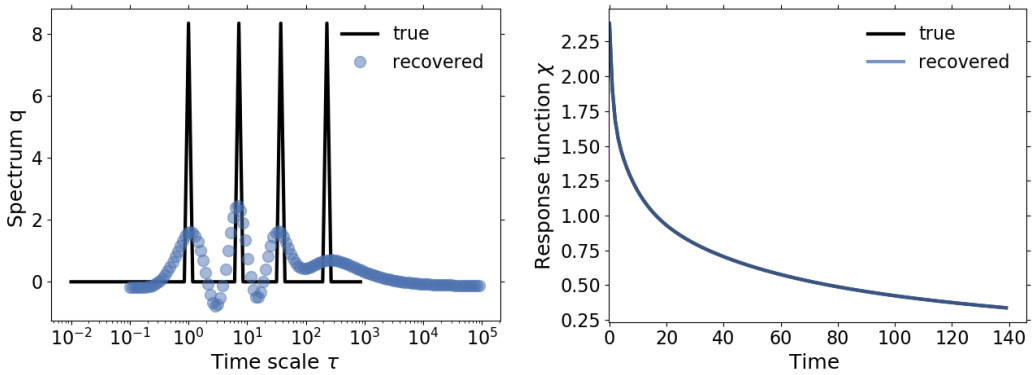


Figure B.2 : (continued) Spectrum $\mathbf{q}_\lambda$ and response function $\chi(t)$ derived for different discrete spectra $\mathbf{q}^*$ for $SNR \sim \mathcal{O}(10^4) - \mathcal{O}(10^5)$. Subfigures (a)-(f) show results for taking data from a 1% experiment and subfigures (g)-(j) for a $2\times f_0$ experiment. The recovered spectra give smooth approximations to the underlying discrete spectra. The response functions are recovered almost perfectly.

(g) $\tau = 1, \tau = 7, \tau = 37, \tau = 100$



(h) $\tau = 1, \tau = 7, \tau = 37, \tau = 228$



(i) $\tau = 1, \tau = 7, \tau = 37, \tau = 228, \tau = 518$

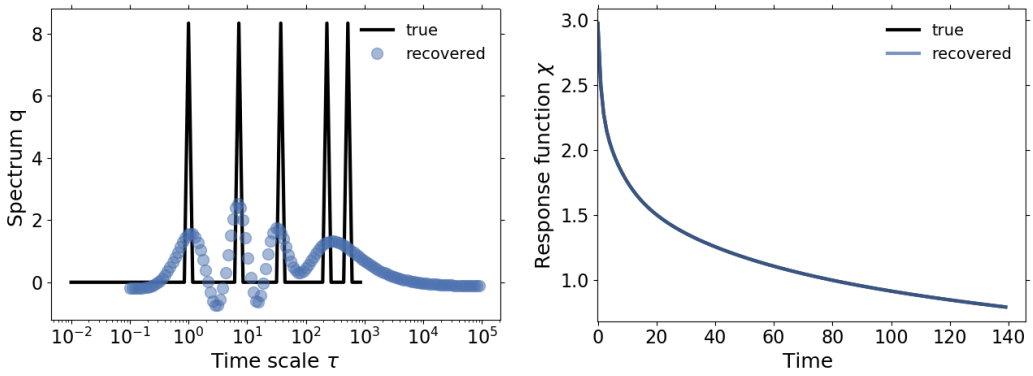


Figure B.2 : (continued) Spectrum $\mathbf{q}_\lambda$ and response function $\chi(t)$ derived for different discrete spectra $\mathbf{q}^*$ for $SNR \sim \mathcal{O}(10^4) - \mathcal{O}(10^5)$. Subfigures (a)-(f) show results for taking data from a 1% experiment and subfigures (g)-(j) for a $2\times f_0$ experiment. The recovered spectra give smooth approximations to the underlying discrete spectra. The response functions are recovered almost perfectly.

C

APPENDIX TO CHAPTER 4

C.1 MONOTONICITY OF $\chi_\gamma^{(O)}$ IN THE MPI-ESM

This appendix is referred to in section 4.3.2 where it is claimed that by enforcing monotonicity with the RFI method (see Fig. 2.2) one obtains for the MPI-ESM a recovery of $\chi_\gamma^{(O)}(t)$ that approximately matches a recovery obtained from an independent method that does not assume monotonicity, which suggests that assuming monotonicity to derive $\chi_\gamma^{(O)}(t)$ is plausible. For the demonstration I compare the recovery of $\chi_\gamma^{(O)}(t)$ from the RFI method enforcing monotonicity with a recovery from the *step method* (Ragone et al., 2016; more details in section 2.5) employing a simplifying assumption to approximate the temperature forcing. Although the recovery by the step method is usually quite "noisy" when data are taken from only one realization of the experiment (see more details in section 2.5), if the signal-to-noise ratio of the response from this realization is sufficiently large this method gives a useful approximation to the response function without assuming monotonicity.

The step method requires the perturbation – in the case of $\chi_\gamma^{(O)}$ the ocean temperature, see eq. (4.3) – to be a Heaviside-type function. Therefore for the recovery with the step method I took data from the $2\times CO_2$ rad experiment (see Table 2.1). Because in this experiment the adjustment of ocean temperature is relatively fast for the first years and slow thereafter (see Fig. C.1a), one can roughly approximate this temperature by a Heaviside-type function so that the step method can be employed. This is equivalent to consider that the ocean temperature response to CO_2 has no memory, which is obviously not the case but can be thought of as a first-order approximation considering the 140 years over which one is interested here (compare also the relatively fast decay of $\chi_\alpha^{(O)}(t)$ that characterizes the ocean temperature response to CO_2 in Fig. 4.3c). To make this approximation I take the mean over the temperature time series ignoring the transient over the first 20 years (see $\Delta T_O^{(2\times CO_2)}$ in Fig. C.1a). After estimating $\Delta T_O^{(2\times CO_2)}$ in this way, an approximation for $\chi_\gamma^{(O)}(t)$ can be obtained from (4.16) by

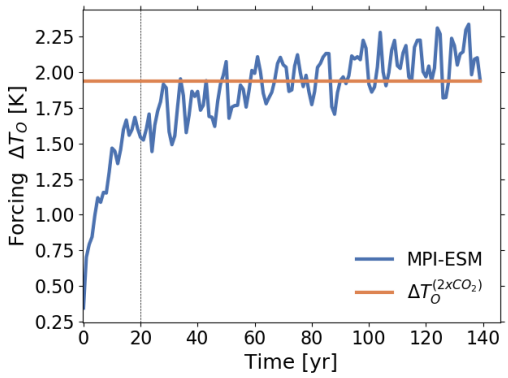
$$\chi_\gamma^{(O)}(t) = \frac{1}{\Delta T_O^{(2\times CO_2)}} \frac{d}{dt} \Delta C_O^{(2\times CO_2)}(t). \tag{C.1}$$

For the recovery with the RFI method I took data from the $2\times CO_2$ and 1% rad experiments and chose to enforce monotonicity (step 6 of Fig. 2.2).

All recoveries are plotted in Fig. C.1b. As seen, the recovery by the step method, although indeed "noisy" and suffering from the error made by approximating $\Delta T_O^{(2\times CO_2)}$, is well fitted by the two recoveries from the completely independent RFI method. Systematic discrepancies are found only at short times, which are anyway harder to recover because of the ill-posedness of (4.12) that obscures high frequencies – and thus small time scales – of the response function (see chapter 2). The discrepancies between the two recoveries by the RFI method are expected because of the difference in the signal-to-noise ratio in the 1% and $2\times CO_2$ rad experiments (see more details in section 3.3).

Despite these discrepancies, the overall agreement of the recoveries shows that the RFI method gives an approximation to $\chi_\alpha^{(O)}(t)$ that is consistent with an approximation from an independent method that does not assume monotonicity, and in particular gives some confidence that assuming monotonicity to derive $\chi_\alpha^{(O)}(t)$ by the RFI method is plausible. More indication of the quality of the recovery by the RFI method is given in section 4.3.2 by showing that one can employ it to predict the response of the model in experiments not used for the recovery. The monotonicity assumption is therefore extended to derive $\chi_\gamma^{(O)}(t)$ also for other CMIP5 models in section 4.5.

(a) Ocean near-surface temperature used as forcing



(b) $\chi_\gamma^{(O)}(t)$ derived from both methods

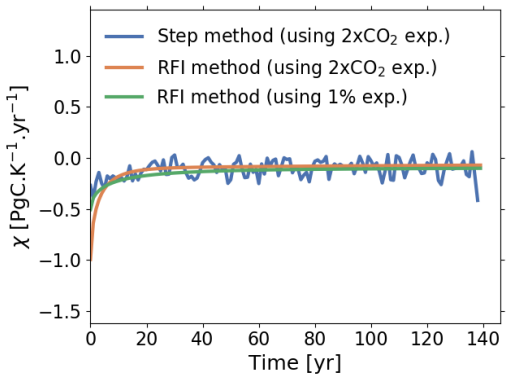


Figure C.1 : Recovery of $\chi_\gamma^{(O)}(t)$ by the RFI method and by the step method (Ragone et al., 2016).

C.2 METHOD TO CALCULATE χ_ζ

In this appendix it is shown in more detail how $\chi_\zeta(t)$ is calculated in section 4.4 enforcing the constraint $\chi_\zeta(0) = 1$. The RFI method recovers a response function $\chi(t)$ employing Tikhonov-Phillips regularization, with the regularization parameter λ determined from an estimate of the noise in the data (see section 2.3.4). But the RFI method does not account for the desired constraint $\chi(0) = 1$. Therefore to derive $\chi_\zeta(t)$ enforcing this constraint I proceed in two steps. First, I derive $\chi_\zeta(t)$ by the RFI method without the constraint, obtaining thereby the value of λ. Then, I use the obtained λ to derive $\chi_\zeta(t)$ employing the same regularization procedure from the RFI method but accounting in addition the constraint $\chi_\zeta(0) = 1$ by the method of Lagrange multipliers (Sundaram, 1996, Selesnick, 2013).

After obtaining λ by the RFI method in the first step, the derivation of $\chi_\zeta(t)$ in the second step proceeds as follows. As in the RFI method, I assume $\chi_\zeta(t)$ to be given by

$$\chi_\zeta(t) = \int_{-\infty}^{\infty} q(\tau) e^{-t/\tau} d\log_{10}\tau, \tag{C.2}$$

where $q(\tau)$ is the spectrum of time scales. By plugging (C.2) into (4.12) and prescribing a distribution of time scales τ, the problem of finding $\chi_\zeta(t)$ boils down to finding the spectrum $q(\tau)$. Once $q(\tau)$ is derived, $\chi_\zeta(t)$ follows from (C.2). For more details please refer to section 2.3.

To understand how the constraint can be enforced, one has to consider

$$\chi_\zeta(0) = \int_{-\infty}^{\infty} q(\tau) d\log_{10}\tau. \tag{C.3}$$

Eq. (C.3) can be discretized assuming a prescribed distribution of time scales equally spaced at a logarithmic scale between maximum and minimum values τ_{max} and τ_{min} (see section 2.3). This discretization gives

$$\chi_\zeta(0) \approx \Delta\log_{10}\tau \sum_{j=0}^{M-1} q_j, \tag{C.4}$$

where M is the number of q_j terms and $\Delta\log_{10}\tau := (\log_{10}\tau_{max} - \log_{10}\tau_{min})/M$ is the spacing between the time scales. The values for all these prescribed parameters are given in section 2.4.2. Using (C.4) one can write the desired constraint $\chi_\zeta(0) = 1$ in a discrete formulation as

$$C\mathbf{q} = 1, \tag{C.5}$$

where C is the row matrix $C := [1, 1, ..., 1]\Delta\log_{10}\tau$.

Knowing how to discretely account for the desired constraint the spectrum $\mathbf{q}$ can now be derived. The procedure consists of minimizing the typical cost function employed in

Tikhonov-Phillips regularization (see e.g. Hansen, 2010, p. 60), which is also done in the RFI method, but subject to the constraint (C.5), i.e.

$$\min_{\mathbf{q}_\lambda} \left(||\mathbf{\Delta Y} - A\mathbf{q}_\lambda||^2 + \lambda ||\mathbf{q}_\lambda||^2 \right) \quad \text{subject to} \quad C\mathbf{q}_\lambda = 1, \tag{C.6}$$

where $||.||$ denotes the Euclidean norm, $\mathbf{q}_\lambda$ is the spectrum vector recovered by regularization, $\mathbf{\Delta Y}$ is the response data vector and A is the matrix obtained after discretization (see eq. (2.17)). The solution can be obtained by the method of Lagrange multipliers:

$$\mathbf{q}_\lambda = (A^T A + \lambda I)^{-1}(A^T \mathbf{\Delta Y} - C^T [C(A^T A + \lambda I)^{-1} C^T]^{-1} [C(A^T A + \lambda I)^{-1} A^T \mathbf{\Delta Y} - 1]), \tag{C.7}$$

where I is the identity matrix. Since λ was already obtained in the first step, eq. (C.7) gives a discrete approximation to the spectrum $q(\tau)$ by which the response function $\chi_\zeta(t)$ is derived employing (C.2).

C.3 $\chi_\zeta(t)$ NON-NEGATIVE AND MONOTONIC IMPLIES $0 \leq \widetilde{A}(p) \leq 1$ AND INTERPRETATION OF A

In this appendix it is shown that if $\chi_\zeta(t)$ is non-negative and monotonic for all times t, then $0 \leq \widetilde{A}(p) \leq 1$ for all time scales $1/p$, as claimed in subsection 4.4.2. It is also demonstrated that the airborne fraction A can be understood as the distribution derivative of χ_ζ considered as distribution, and thereby as the impulse response of atmospheric CO_2 flux to the CO_2 emissions flux, as claimed in subsection 4.5.2.

To prepare for the demonstration, one must recognize the implicit assumption in the linear response framework that the linear response functions χ are causal, i.e. $\chi(t) = 0$ for $t < 0$. An additional assumption will be that $\chi_\zeta(t) \to 0$ as $t \to \infty$, which is in agreement with results from simulation of the global carbon cycle (e.g. Joos et al., 2013). Because in the present application p is interpreted as a rate and $1/p$ as a time scale, the analysis will focus only on $p > 0$ or, equivalently, on $1/p > 0$. Since $\chi_\zeta(t) \to 0$ as $t \to \infty$ and I only consider $p > 0$, problems with the convergence of the following Laplace transforms are prevented.

Since χ_ζ is assumed to be causal but has the constraint $\chi_\zeta(0) = 1$ (see section 4.4), it has a jump discontinuity at $t = 0$. Because in the following it will be necessary to compute the time derivative of χ_ζ, which due to the discontinuity is not defined at $t = 0$, I will consider the concept of distributions or generalized functions (Schwartz, 1966, Doetsch, 1974, Beerends et al., 2003). Following Beerends et al., 2003 (p. 201), let χ_ζ be a continuously differentiable function for $t \neq 0$ with a finite jump discontinuity at $t = 0$. Let χ'_ζ denote the first time derivative of χ_ζ at all times $t \neq 0$. Assume that χ_ζ and χ'_ζ can be considered as distributions T_χ and $T_{\chi'}$ (see eq. (8.12) in Beerends et al., 2003). Then the distribution derivative T'_χ is given by the jump formula (eq. (8.21) in Beerends et al., 2003 or more generally remark 20. in Doetsch, 1974, p. 315)

$$T'_\chi = T_{\chi'} + \left(\chi_\zeta(0^+) - \chi_\zeta(0^-) \right) \delta, \tag{C.8}$$

where $\chi_\zeta(0^+)$ and $\chi_\zeta(0^-)$ are the limits of $\chi_\zeta(t)$ for $t \to 0$ taken from the right and from the left, and δ is the Dirac delta distribution. Because χ_ζ is assumed to be causal,

$$\chi_\zeta(0^-) = 0. \tag{C.9}$$

Moreover $\chi_\zeta(0) = 1$, so that

$$\chi_\zeta(0^+) = 1. \tag{C.10}$$

Therefore (C.8) reads

$$T'_\chi = T_{\chi'} + \delta. \tag{C.11}$$

Applying a Laplace transform to (C.11) gives (Beerends et al., 2003, Table 10)

$$\mathcal{L}\{T'_\chi\}(p) = \mathcal{L}\{T_{\chi'}\}(p) + \mathcal{L}\{\delta\}(p). \tag{C.12}$$

It follows from the differentiation property of the Laplace transform of distributions (Beerends et al., 2003, Table 10; Doetsch, 1974, p. 317) that

$$\mathcal{L}\{T'_\chi\}(p) = p\mathcal{L}\{T_\chi\}. \tag{C.13}$$

But the Laplace transform of the distribution T_χ is equal to the Laplace transform of χ_ζ (Beerends et al., 2003, eq. (13.8); Doetsch, 1974, Theorem 12.2), i.e.

$$\mathcal{L}\{T_\chi\}(p) = \widetilde{\chi}_\zeta(p). \tag{C.14}$$

Therefore (C.13) gives

$$\mathcal{L}\{T'_\chi\}(p) = p\widetilde{\chi}_\zeta(p) \tag{C.15}$$

and from eq. (4.21) it follows that

$$\mathcal{L}\{T'_\chi\}(p) = \widetilde{A}(p), \tag{C.16}$$

so that the airborne fraction A can be regarded as the distribution derivative of χ_ζ considered as a distribution T_χ. Because χ_ζ can be interpreted as the impulse response of atmospheric CO_2 to the CO_2 emissions flux, its distribution derivative and by eq. (C.16) also the airborne fraction A can be interpreted as the impulse response of atmospheric CO_2 flux to the CO_2 emissions flux. This is the argument used in subsection 4.5.2 to interpret the behaviour of $\widetilde{A}(p)$ by its correspondence to the impulse response of the atmospheric CO_2 flux in the time domain.

In addition, the Laplace transform of $T_{\chi'}$ is equal to the Laplace transform of χ'_ζ, i.e.

$$\mathcal{L}\{T_{\chi'}\}(p) = \widetilde{\chi'_\zeta}(p), \tag{C.17}$$

and the Laplace transform of δ is

$$\mathcal{L}\{\delta\}(p) = 1. \tag{C.18}$$

Hence, plugging eqs. (C.16), (C.17), (C.18) into (C.12) gives

$$\widetilde{A}(p) = \widetilde{\chi'_\zeta}(p) + 1. \tag{C.19}$$

The Laplace transform $\widetilde{\chi'_\zeta}(p)$ may be written as (Lundberg et al., 2007)

$$\widetilde{\chi'_\zeta}(p) = \int_{0^+}^{\infty} \chi'_\zeta(t) e^{-pt} dt. \tag{C.20}$$

Therefore, if $\chi_\zeta(t)$ is non-negative and monotonically decreasing for all $t > 0$, then $\chi'_\zeta(t) \leq 0$ for all $t > 0$ (Theorem 6.2.7 in Bartle and Sherbert, 2011). Because $e^{-pt} \geq 0$ for all $t, p > 0$,

$$\widetilde{\chi'_\zeta}(p) \leq 0 \quad \text{for all} \quad p > 0, \tag{C.21}$$

which implies

$$\widetilde{\chi'_\zeta}(p) + 1 \leq 1 \quad \text{for all} \quad p > 0. \tag{C.22}$$

But from eq. (C.19) the left-hand side of (C.22) equals $\widetilde{A}(p)$. Hence

$$\widetilde{A}(p) \leq 1 \quad \text{for all} \quad p > 0. \tag{C.23}$$

To determine the lower bound of $\widetilde{A}(p)$, eq. (4.21) can be rewritten as

$$\widetilde{A}(p) = p \int_{0^+}^{\infty} \chi_\zeta(t) e^{-pt} dt. \tag{C.24}$$

If $\chi_\zeta(t)$ is non-negative for all $t > 0$, because $e^{-pt} \geq 0$ for all $t, p > 0$, it follows that $\widetilde{\chi}_\zeta(p) \geq 0$ for all $p > 0$. Hence

$$p\widetilde{\chi}_\zeta(p) = \widetilde{A}(p) \geq 0 \quad \text{for all} \quad p > 0. \tag{C.25}$$

Because the time scales $1/p$ are merely a way of presenting the domain of $\widetilde{A}(p)$, (C.23) and (C.25) are valid also for all time scales $1/p > 0$.

Since $\chi_\zeta(t)$ has been derived in section 4.4, $\widetilde{\chi'_\zeta}(p)$ can be computed so that the decomposition of $\widetilde{A}(p)$ given by eq. (C.19) can be numerically checked. Fig. C.2 compares this decomposition to the airborne fraction $\widetilde{A}(p)$ obtained in Fig. 4.6 from the right-hand side of eq. (4.21). As seen, indeed the results match perfectly.

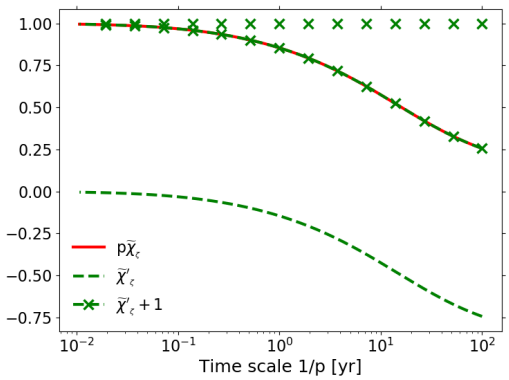


Figure C.2 : Decomposition of airborne fraction given by eq. (C.19) compared against eq. (4.21).

BIBLIOGRAPHY

Adloff, M., C. H. Reick, and M. Claussen (2018). "Earth system model simulations show different feedback strengths of the terrestrial carbon cycle under glacial and interglacial conditions." *Earth System Dynamics* 9, pp. 413–425 (cit. on pp. 4, 75, 78, 79).

Aengenheyster, M., Q. Y. Feng, F. Van Der Ploeg, and H. A. Dijkstra (2018). "The point of no return for climate action: effects of climate uncertainty and risk tolerance." *Earth System Dynamics* 9.3 (cit. on pp. 5, 11, 40, 125).

Alexandrov, G., T Oikawa, and Y Yamagata (2003). "Climate dependence of the CO_2 fertilization effect on terrestrial net primary production." *Tellus B: Chemical and Physical Meteorology* 55.2, pp. 669–675 (cit. on p. 57).

Anderson, T. R., E. Hawkins, and P. D. Jones (2016). "CO_2, the greenhouse effect and global warming: from the pioneering work of Arrhenius and Callendar to today's Earth System Models." *Endeavour* 40.3, pp. 178–187 (cit. on p. 1).

Anderssen, R. S. and P. Bloomfield (1974). "Numerical differentiation procedures for non-exact data." *Numerische Mathematik* 22.3, pp. 157–182 (cit. on p. 40).

Archer, D. (2010). *The global carbon cycle*. Princeton University Press (cit. on p. 3).

Arneth, A. et al. (2010). "Terrestrial biogeochemical feedbacks in the climate system." *Nature Geoscience* 3.8, pp. 525–532 (cit. on pp. 4, 75).

Arora, V. K. et al. (2013). "Carbon–concentration and carbon–climate feedbacks in CMIP5 Earth system models." *Journal of Climate* 26.15, pp. 5289–5314 (cit. on pp. 1, 2, 4, 9, 73, 75–77, 79, 100–103, 107, 108, 112).

Arora, V. K. et al. (2019). "Carbon-concentration and carbon-climate feedbacks in CMIP6 models, and their comparison to CMIP5 models." *Biogeosciences Discussions*, pp. 1–124 (cit. on p. 1).

Arrhenius, S. (1896). "On the influence of carbonic acid in the air upon the temperature of the ground." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 41.251, pp. 237–276 (cit. on p. 1).

Bacastow, R. B. and C. D. Keeling (1979). "Models to predict future atmospheric CO2 concentrations." In: *Workshop on the global effects of carbon dioxide from fossil fuels*. US Department of Energy, pp. 72–90 (cit. on p. 2).

Bakushinskii, A. (1984). "Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion." *USSR Computational Mathematics and Mathematical Physics* 24.4, pp. 181–182 (cit. on p. 46).

Bartle, R. and D. Sherbert (2011). *Introduction to Real Analysis, 4th Edition*. John Wiley & Sons, Incorporated. ISBN: 9781118135860. URL: https://books.google.de/books?id=YawbAAAAQBAJ (cit. on p. 153).

Beerends, R. J., H. G. ter Morsche, J. Van den Berg, and E. Van de Vrie (2003). *Fourier and Laplace transforms* (cit. on pp. 97, 98, 105, 151, 152).

Bennedsen, M., E. Hillebrand, and S. J. Koopman (2019). "Trend analysis of the airborne fraction and sink rate of anthropogenically released CO2." *Biogeosciences* 16.18, pp. 3651–3663 (cit. on p. 1).

Bertero, M. (1989). "Linear inverse and ill-posed problems." In: *Advances in Electronics and Electron Physics*. Vol. 75. Elsevier, pp. 1–120 (cit. on pp. 12, 20, 21).

Bertero, M., P. Boccacci, and F Maggio (1995). "Regularization methods in image restoration: an application to HST images." *International Journal of Imaging Systems and Technology* 6.4, pp. 376–386 (cit. on pp. 12, 21, 43).

Bódai, T., V. Lucarini, and F. Lunkeit (2020). "Can we use linear response theory to assess geoengineering strategies?" *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30.2, p. 023124 (cit. on pp. 5, 11, 16, 125).

Boer, G. and V. Arora (2013). "Feedbacks in emission-driven and concentration-driven global carbon budgets." *Journal of climate* 26.10, pp. 3326–3341 (cit. on pp. 1, 4, 75).

Canadell, J. G. et al. (2007). "Contributions to accelerating atmospheric CO2 growth from economic activity, carbon intensity, and efficiency of natural sinks." *Proceedings of the national academy of sciences* 104.47, pp. 18866–18870 (cit. on p. 97).

Carlson, C., N. Bates, D. Hansell, and D. Steinberg (2001). "Carbon Cycle." English (US). In: *Encyclopedia of Ocean Sciences*. Elsevier Inc., pp. 477–486. ISBN: 9780123744739. DOI: 10.1016/B978-012374473-9.00272-1 (cit. on p. 2).

Ciais, P. et al. (2013). "Carbon and Other Biogeochemical Cycles." 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*. Ed. by T. Stocker et al. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, pp. 465–570. ISBN: ISBN 978-1-107-66182-0 (cit. on pp. 2, 3, 18, 78, 114).

Colbourn, G, A Ridgwell, and T. Lenton (2015). "The time scale of the silicate weathering negative feedback on atmospheric CO_2." *Global Biogeochemical Cycles* 29.5, pp. 583–596 (cit. on pp. 17, 18).

Collatz, G. J., M Ribas-Carbo, and J. Berry (1992). "Coupled photosynthesis-stomatal conductance model for leaves of C4 plants." *Functional Plant Biology* 19.5, pp. 519–538 (cit. on p. 127).

Cox, P. M., C. Huntingford, and M. S. Williamson (2018). "Emergent constraint on equilibrium climate sensitivity from global temperature variability." *Nature* 553.7688, pp. 319–322 (cit. on pp. 10, 125).

Curry, J. A., J. L. Schramm, and E. E. Ebert (1995). "Sea ice-albedo climate feedback mechanism." *Journal of Climate* 8.2, pp. 240–247 (cit. on p. 3).

Doetsch, G. (1974). *Introduction to the Theory and Application of the Laplace Transformation*. Springer-Verlag. ISBN: 9780387064079 (cit. on pp. 97, 151, 152).

Dunne, J. P. et al. (2013). "GFDL's ESM2 global coupled climate–carbon earth system models. Part II: carbon system formulation and baseline simulation characteristics." *Journal of Climate* 26.7, pp. 2247–2267 (cit. on p. 76).

Emanuel, W. R., G. E. Killough, and J. S. Olson (1981). "Modelling the Circulation of Carbon in the World's Terrestrial Ecosystems." *SCOPE 16*. Ed. by B. Bolin, pp. 335–353 (cit. on p. 11).

Engl, H. W., M. Hanke, and A. Neubauer (1996). *Regularization of inverse problems*. Vol. 375. Springer Science & Business Media (cit. on pp. 12, 18, 21, 22, 40, 46, 47).

Enting, I. and N. Clisby (2019). "Estimates of climatic influence on the carbon cycle." *Earth System Dynamics Discussions* 2019, pp. 1–21. DOI: 10.5194/esd-2019-41. URL: https://www.earth-syst-dynam-discuss.net/esd-2019-41/ (cit. on pp. 5, 9, 11, 13, 75, 76, 95, 110, 114).

Enting, I. (1990). "Ambiguities in the calibration of carbon cycle models." *Inverse Problems* 6.5, p. L39 (cit. on p. 11).

Enting, I. (2007). "Laplace transform analysis of the carbon cycle." *Environmental Modelling & Software* 22.10, pp. 1488–1497 (cit. on pp. 105, 114).

Enting, I. and J. Mansbridge (1987). "Inversion relations for the deconvolution of CO_2 data from ice cores." *Inverse Problems* 3.4, p. L63 (cit. on p. 17).

Farquhar, G. D., S. v. von Caemmerer, and J. A. Berry (1980). "A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species." *Planta* 149.1, pp. 78–90 (cit. on p. 127).

Field, C. B. and M. R. Raupach (2004). *The global carbon cycle: integrating humans, climate, and the natural world*. Vol. 62. Island Press (cit. on pp. 3, 4).

Forney, D. and D. Rothman (2012a). "Inverse method for estimating respiration rates from decay time series." *Biogeosciences* 9.9, pp. 3601–3612 (cit. on pp. 18, 28).

Forney, D. C. and D. H. Rothman (2012b). "Common structure in the heterogeneity of plant-matter decay." *Journal of the Royal Society Interface* 9.74, pp. 2255–2267 (cit. on pp. 48, 125).

Friedlingstein, P, J.-L. Dufresne, P. Cox, and P Rayner (2003). "How positive is the feedback between climate change and the carbon cycle?" *Tellus B* 55.2, pp. 692–700 (cit. on pp. 4, 9, 13, 75–77, 79, 110).

Friedlingstein, P. (2015). "Carbon cycle feedbacks and future climate change." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 373.2054, p. 20140421 (cit. on pp. 1, 3, 4).

Friedlingstein, P. et al. (2006). "Climate–carbon cycle feedback analysis: results from the C4MIP model intercomparison." *Journal of climate* 19.14, pp. 3337–3353 (cit. on pp. 1, 4, 9, 75, 100, 107).

Friedlingstein, P. et al. (2014). "Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks." *Journal of Climate* 27.2, pp. 511–526 (cit. on pp. 4, 75, 108).

Friedlingstein, P. et al. (2019). "Global carbon budget 2019." *Earth System Science Data* 11.4, pp. 1783–1838 (cit. on pp. 1, 3).

Fung, I, P. Rayner, P Friedlingstein, and D Sahagian (2000). "Full-form Earth System models: coupled carbon-climate interaction experiment (the'Flying Leap')." *IGBP Global Change Newsletter* 41 (cit. on p. 9).

Gallavotti, G. (2014). *Nonequilibrium and irreversibility*. Springer (cit. on p. 15).

Gallavotti, G. and E. G. D. Cohen (1995). "Dynamical ensembles in stationary states." *Journal of Statistical Physics* 80.5-6, pp. 931–970 (cit. on p. 10).

Ghil, M. and V. Lucarini (2019). "The Physics of Climate Variability and Climate Change." *arXiv preprint arXiv:1910.00583* (cit. on pp. 11, 125).

Giorgetta, M. A. et al. (2013). "Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5." *Journal of Advances in Modeling Earth Systems* 5.3, pp. 572–597 (cit. on p. 127).

Gloor, M., J. L. Sarmiento, and N. Gruber (2010). "What can be learned about carbon cycle climate feedbacks from the CO2 airborne fraction?" *Atmospheric Chemistry and Physics* 10.16, pp. 7739–7751 (cit. on p. 97).

Golub, G. H. and C. F. Van Loan (1996). *Matrix Computations*. The Johns Hopkins University Press (cit. on p. 20).

Goodwin, P., R. G. Williams, V. M. Roussenov, and A. Katavouta (2019). "Climate sensitivity from both physical and carbon cycle feedbacks." *Geophysical Research Letters* 46.13, pp. 7554–7564 (cit. on pp. 4, 75, 114, 124).

Gottwald, G. A. (2020). *Introduction to Focus Issue: Linear response theory: Potentials and limits* (cit. on pp. 5, 125).

Gregory, J. M. et al. (2004). "A new method for diagnosing radiative forcing and climate sensitivity." *Geophysical research letters* 31.3 (cit. on p. 64).

Gregory, J. M., C. Jones, P Cadule, and P Friedlingstein (2009). "Quantifying carbon cycle feedbacks." *Journal of Climate* 22.19, pp. 5232–5250 (cit. on pp. 2, 4, 5, 9, 73, 75, 77, 78, 97, 107, 108, 112, 114, 124).

Grieser, J. and C.-D. Schönwiese (2001). "Process, Forcing, and Signal Analysis of Global Mean Temperature Variations by Means of a Three-Box Energy Balance Model." *Climatic Change* 48.4, pp. 617–646 (cit. on p. 17).

Groetsch, C. (1983). "Comments on Morozov's discrepancy principle." In: *Improperly posed problems and their numerical treatment*. Springer, pp. 97–104 (cit. on p. 22).

Groetsch, C. (1984). "The theory of Tikhonov regularization for Fredholm equations." *Boston Pitman Publication* (cit. on pp. 12, 18, 20–22, 24, 33, 128, 141, 142).

Groetsch, C. W. (2003). "The delayed emergence of regularization theory." *Bollettino di storia delle scienze matematiche* 23, pp. 105–120 (cit. on p. 21).

Hämarik, U., R. Palm, and T. Raus (2011). "Comparison of parameter choices in regularization algorithms in case of different information about noise level." *Calcolo* 48.1, pp. 47–59 (cit. on p. 22).

Hansen, P. C. (1987). "The truncated SVD as a method for regularization." *BIT Numerical Mathematics* 27.4, pp. 534–553 (cit. on p. 21).

Hansen, P. C. (1989). "Regularization, GSVD and truncated GSVD." *BIT numerical mathematics* 29.3, pp. 491–504 (cit. on p. 20).

Hansen, P. C. (1990). "The discrete Picard condition for discrete ill-posed problems." *BIT Numerical Mathematics* 30.4, pp. 658–672 (cit. on pp. 22, 46).

Hansen, P. C. (1992). "Numerical tools for analysis and solution of Fredholm integral equations of the first kind." *Inverse problems* 8.6, p. 849 (cit. on pp. 20, 31).

Hansen, P. C. (2002). "Deconvolution and regularization with Toeplitz matrices." *Numerical Algorithms* 29.4, pp. 323–378 (cit. on pp. 72, 134).

Hansen, P. C. (2010). *Discrete inverse problems: insight and algorithms*. Vol. 7. Siam (cit. on pp. 12, 21, 22, 151).

Hasselmann, K., R. Sausen, E. Maier-Reimer, and R. Voss (1993). "On the cold start problem in transient simulations with coupled atmosphere-ocean models." *Climate Dynamics* 9.2, pp. 53–61 (cit. on pp. 11, 17, 18, 40).

Hasselmann, K., S. Hasselmann, R. Giering, V. Ocana, and H. Storch (1997). "Sensitivity study of optimal CO_2 emission paths using a simplified structural integrated assessment model (SIAM)." *Climatic Change* 37.2, pp. 345–386 (cit. on p. 17).

Heinze, C. et al. (2015). "The ocean carbon sink–impacts, vulnerabilities and challenges." *Earth System Dynamics* 6.1, pp. 327–358 (cit. on p. 3).

Hooß, G., R. Voss, K. Hasselmann, E. Maier-Reimer, and F. Joos (2001). "A nonlinear impulse response model of the coupled carbon cycle-climate system (NICCS)." *Climate Dynamics* 18.3-4, pp. 189–202 (cit. on p. 81).

Houghton, R. A. (2003). "The contemporary carbon cycle." *Treatise on geochemistry* 8, pp. 473–513 (cit. on pp. 2, 3).

IPCC (2013). "Summary for Policymakers." 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*. Ed. by T. Stocker et al. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Chap. SPM, pp. 1–30. ISBN: ISBN 978-1-107-66182-0. DOI: 10.1017/CBO9781107415324.004. URL: www.climatechange2013.org (cit. on p. 1).

Ilyina, T. and P. Friedlingstein (2016). "Biogeochemical cycles and climate change." *White paper on WCRP Grand Challenge (World Climate Research Programme)* (cit. on pp. 1, 3).

Ilyina, T. et al. (2013). "Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations." *Journal of Advances in Modeling Earth Systems* 5.2, pp. 287–315 (cit. on p. 127).

Istratov, A. A. and O. F. Vyvenko (1999). "Exponential analysis in physical phenomena." *Review of Scientific Instruments* 70.2, pp. 1233–1257 (cit. on pp. 12, 18, 43, 48, 72).

Jones, C. D. and P. Friedlingstein (2020). "Quantifying process-level uncertainty contributions to TCRE and Carbon Budgets for meeting Paris Agreement climate targets." *Environmental Research Letters* (cit. on pp. 2, 4, 75, 108, 113, 114, 123, 124).

Jones, C., J. Lowe, S. Liddicoat, and R. Betts (2009). "Committed terrestrial ecosystem changes due to climate change." *Nature Geoscience* 2.7, pp. 484–487 (cit. on p. 10).

Jones, C. et al. (2013). "Twenty-first-century compatible CO2 emissions and airborne fraction simulated by CMIP5 earth system models under four representative concentration pathways." *Journal of Climate* 26.13, pp. 4398–4413 (cit. on pp. 1, 97).

Joos, F and M Bruno (1996). "Pulse response functions are cost-efficient tools to model the link between carbon emissions, atmospheric CO_2 and global warming." *Physics and Chemistry of the Earth* 21.5-6, pp. 471–476 (cit. on pp. 11, 40, 81).

Joos, F. et al. (1996). "An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake." *Tellus B* 48.3, pp. 397–417 (cit. on pp. 11, 40).

Joos, F. et al. (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, pp. 2793–2825 (cit. on pp. 2, 11, 17, 40, 95, 151).

Jungclaus, J. et al. (2013). "Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model." *Journal of Advances in Modeling Earth Systems* 5.2, pp. 422–446 (cit. on p. 127).

Keppler, L. (2020). "Variability of the contemporary Southern Ocean carbon fluxes and storage." PhD thesis. Universität Hamburg Hamburg (cit. on p. 3).

Kubo, R. (1957). "Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems." *Journal of the Physical Society of Japan* 12.6, pp. 570–586 (cit. on pp. 5, 10).

Kubo, R., M. Toda, and N. Hashitsume (2012). *Statistical physics II: nonequilibrium statistical mechanics*. Vol. 31. Springer Science & Business Media (cit. on p. 5).

Kumaresan, R., D. Tufts, and L. L. Scharf (1984). "A Prony method for noisy data: Choosing the signal components and selecting the order in exponential signal models." *Proceedings of the IEEE* 72.2, pp. 230–233 (cit. on p. 18).

Lanczos, C. (1956). *Applied Analysis*. Prentice Hall (cit. on pp. 18, 72).

Le Quéré, C. et al. (2009). "Trends in the sources and sinks of carbon dioxide." *Nature geoscience* 2.12, pp. 831–836 (cit. on p. 1).

Lembo, V., V. Lucarini, and F. Ragone (2020). "Beyond forcing scenarios: predicting climate change through response operators in a coupled general circulation model." *Scientific Reports* 10.1, pp. 1–13 (cit. on pp. 5, 10, 11, 95, 125).

Levin, E. and A. Y. Meltzer (2017). "Estimation of the regularization parameter in linear discrete ill-posed problems using the Picard parameter." *SIAM Journal on Scientific Computing* 39.6, A2741–A2762 (cit. on p. 46).

Li, S. and A. Jarvis (2009). "Long run surface temperature dynamics of an A-OGCM: the HadCM3 4× CO_2 forcing experiment revisited." *Climate Dynamics* 33.6, pp. 817–825 (cit. on p. 17).

Lord, N. S., A Ridgwell, M. Thorne, and D. Lunt (2016). "An impulse response function for the "long tail" of excess atmospheric CO_2 in an Earth system model." *Global Biogeochemical Cycles* 30.1, pp. 2–17 (cit. on pp. 17, 18).

Lucarini, V. (2008). "Response theory for equilibrium and non-equilibrium statistical mechanics: Causality and generalized Kramers-Kronig relations." *Journal of Statistical Physics* 131.3, pp. 543–558 (cit. on p. 15).

Lucarini, V. (2009). "Evidence of dispersion relations for the nonlinear response of the Lorenz 63 system." *Journal of Statistical Physics* 134.2, pp. 381–400 (cit. on pp. 10, 11, 15, 73, 125, 126).

Lucarini, V. and S. Sarno (2010). "A statistical mechanical approach for the computation of the climatic response to general forcings." *arXiv preprint arXiv:1008.0340* (cit. on pp. 5, 10, 15, 125).

Lucarini, V. et al. (2014). "Mathematical and physical ideas for climate science." *Reviews of Geophysics* 52.4, pp. 809–859. DOI: 10.1002/2013RG000446. eprint: https:

//agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013RG000446 (cit. on pp. 10, 11, 125).

Lucarini, V., F. Ragone, and F. Lunkeit (2017). "Predicting climate change using response theory: Global averages and spatial patterns." *Journal of Statistical Physics* 166.3-4, pp. 1036–1064 (cit. on pp. 5, 11, 12, 15, 16, 40, 125).

Lundberg, K. H., H. R. Miller, and D. L. Trumper (2007). "Initial conditions, generalized functions, and the laplace transform troubles at the origin." *IEEE Control Systems Magazine* 27.1, pp. 22–35 (cit. on p. 153).

MacDougall, A. H. et al. (2020). "Is there warming in the pipeline? A multi-model analysis of the Zero Emissions Commitment from CO_2." *Biogeosciences* 17, pp. 2987–3016 (cit. on pp. 114, 124).

MacMartin, D. G. and B. Kravitz (2016). "Dynamic climate emulators for solar geoengineering." *Atmospheric Chemistry and Physics* 16.24, pp. 15789–15799 (cit. on pp. 11, 12, 40, 43, 53).

Mackenzie, F. T. and A. Lerman (2006). *Carbon in the Geobiosphere:-Earth's Outer Shell*. Vol. 25. Springer Science & Business Media (cit. on p. 3).

Maier-Reimer, E. and K. Hasselmann (1987). "Transport and storage of CO_2 in the ocean - an inorganic ocean-circulation carbon cycle model." *Climate dynamics* 2.2, pp. 63–90 (cit. on pp. 11, 17, 18, 40).

Marotzke, J. et al. (2017). "Climate research must sharpen its view." *Nature climate change* 7.2, pp. 89–91 (cit. on p. 1).

Mauritsen, T. and R. Pincus (2017). "Committed warming inferred from observations." *Nature Climate Change* 7.9, p. 652 (cit. on pp. 10, 114, 124).

Melillo, J. et al. (2002). "Soil warming and carbon-cycle feedbacks to the climate system." *Science* 298.5601, pp. 2173–2176 (cit. on p. 102).

Morozov, V. A. (1966). "On the solution of functional equations by the method of regularization." In: *Doklady Akademii Nauk*. Vol. 167. Russian Academy of Sciences, pp. 510–512 (cit. on p. 22).

Myhre, G. et al. (2013). "Anthropogenic and Natural Radiative Forcing." 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*. Ed. by T. Stocker et al. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. Chap. 8, pp. 659–740. ISBN: ISBN 978-1-107-66182-0. DOI: 10.1017/CBO9781107415324.018. URL: www.climatechange2013.org (cit. on p. 2).

Myhre, G., E. J. Highwood, K. P. Shine, and F. Stordal (1998). "New estimates of radiative forcing due to well mixed greenhouse gases." *Geophysical research letters* 25.14, pp. 2715–2718 (cit. on p. 87).

Nijsse, F. J. and H. A. Dijkstra (2018). "A mathematical approach to understanding emergent constraints." *Earth System Dynamics* 9.3, pp. 999–1012 (cit. on p. 10).

Palm, R. (2010). "Numerical comparison of regularization algorithms for solving ill-posed problems." PhD thesis. Tartu University Press (cit. on p. 21).

Phillips, D. L. (1962). "A technique for the numerical solution of certain integral equations of the first kind." *Journal of the ACM (JACM)* 9.1, pp. 84–97 (cit. on p. 21).

Plattner, G.-K. et al. (2008). "Long-term climate commitments projected with climate–carbon cycle models." *Journal of Climate* 21.12, pp. 2721–2751 (cit. on pp. 10, 114, 124).

Poincaré, H. (1908). *Science and method* (cit. on p. V).

Pongratz, J., K. Caldeira, C. H. Reick, and M. Claussen (2011). "Coupled climate–carbon simulations indicate minor global effects of wars and epidemics on atmospheric CO_2 between ad 800 and 1850." *The Holocene* 21.5, pp. 843–851 (cit. on pp. 11, 18).

Post, W. M. et al. (1990). "The global carbon cycle." *American scientist* 78.4, pp. 310–326 (cit. on p. 2).

Prentice, I. C. et al. (2001). "The carbon cycle and atmospheric carbon dioxide." In: Cambridge University Press (cit. on pp. 2, 3).

Ragone, F., V. Lucarini, and F. Lunkeit (2016). "A new framework for climate sensitivity and prediction: a modelling perspective." *Climate dynamics* 46.5-6, pp. 1459–1471 (cit. on pp. 5, 10–12, 15–17, 40, 74, 124, 125, 147, 149).

Raich, J. W. and C. S. Potter (1995). "Global patterns of carbon dioxide emissions from soils." *Global biogeochemical cycles* 9.1, pp. 23–36 (cit. on pp. 4, 66, 75).

Raupach, M. (2013). "The exponential eigenmodes of the carbon-climate system, and their implications for ratios of responses to forcings." *Earth System Dynamics* 4.1 (cit. on pp. 2, 114).

Raupach, M., J. Canadell, and C. Le Quéré (2008). "Anthropogenic and biophysical contributions to increasing atmospheric CO2 growth rate and airborne fraction." *Biogeosciences* 5.6 (cit. on p. 97).

Raupach, M. R. et al. (2014). "The declining uptake rate of atmospheric CO2 by land and ocean sinks." *Biogeosciences* 11.13, pp. 3453–3475 (cit. on p. 1).

Reick, C., T. Raddatz, V. Brovkin, and V. Gayler (2013). "Representation of natural and anthropogenic land cover change in MPI-ESM." *Journal of Advances in Modeling Earth Systems* 5.3, pp. 459–482 (cit. on p. 127).

Reick, C. H. (2002). "Linear response of the Lorenz system." *Physical Review E* 66.3, p. 036103 (cit. on pp. 11, 15).

Roe, G. (2009). "Feedbacks, timescales, and seeing red." *Annual Review of Earth and Planetary Sciences* 37, pp. 93–115 (cit. on pp. 78, 79, 108, 126).

Rubino, M et al. (2016). "Low atmospheric CO_2 levels during the Little Ice Age due to cooling-induced terrestrial uptake." *Nature Geoscience* 9.9, pp. 691–694 (cit. on pp. 5, 9, 13, 75, 76, 110).

Ruelle, D. (1997). "Differentiation of SRB states." *Communications in Mathematical Physics* 187.1, pp. 227–241 (cit. on p. 10).

Ruelle, D. (1998a). "General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium." *Physics Letters A* 245.3-4, pp. 220–224 (cit. on p. 10).

Ruelle, D. (1998b). "Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms." *Nonlinearity* 11.1, p. 5 (cit. on pp. 73, 126).

Ruelle, D. (1999). "Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics." *Journal of Statistical Physics* 95.1-2, pp. 393–468 (cit. on p. 15).

Ruelle, D. (2009). "A review of linear response theory for general differentiable dynamical systems." *Nonlinearity* 22.4, p. 855 (cit. on p. 5).

Sarmiento, J. L. and M. Bender (1994). "Carbon biogeochemistry and climate change." *Photosynthesis Research* 39.3, pp. 209–234 (cit. on p. 3).

Schlesinger, W. H. (1991). "The Terrestrial Biosphere." In: *Biogeochemistry: an Analysis of Global Change*. Elsevier, pp. 108–141. DOI: 10.1016/b978-0-12-625157-9.50010-1 (cit. on p. 2).

Schneck, R., C. H. Reick, and T. Raddatz (2013). "Land contribution to natural CO_2 variability on time scales of centuries." *Journal of Advances in Modeling Earth Systems* 5.2, pp. 354–365 (cit. on p. 127).

Schwartz, L. (1966). *Théorie des distributions*. Vol. 2. Hermann Paris (cit. on p. 151).

Schwinger, J. et al. (2014). "Nonlinearity of ocean carbon cycle feedbacks in CMIP5 earth system models." *Journal of Climate* 27.11, pp. 3869–3888 (cit. on pp. 1, 4, 75).

Selesnick, I. (2013). "Least squares with examples in signal processing." *Connexions* 4 (cit. on p. 150).

Siegenthaler, U. and H. Oeschger (1978). "Predicting future atmospheric carbon dioxide levels." *Science* 199.4327, pp. 388–395 (cit. on pp. 11, 40).

Singhal, G., G Renger, S. Sopory, K. Irrgang, et al. (2012). *Concepts in photobiology: photosynthesis and photomorphogenesis*. Springer Science & Business Media (cit. on p. 2).

Stevens, B. et al. (2013). "Atmospheric component of the MPI-M Earth system model: ECHAM6." *Journal of Advances in Modeling Earth Systems* 5.2, pp. 146–172 (cit. on p. 127).

Sundaram, R. K. et al. (1996). *A first course in optimization theory*. Cambridge university press (cit. on p. 150).

Taroudaki, V. and D. P. O'Leary (2015). "Near-optimal spectral filtering and error estimation for solving ill-posed problems." *SIAM Journal on Scientific Computing* 37.6, A2947–A2968 (cit. on p. 46).

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012). "An overview of CMIP5 and the experiment design." *Bulletin of the American Meteorological Society* 93.4, pp. 485–498 (cit. on pp. 9, 76, 95).

Thompson, M. V. and J. T. Randerson (1999). "Impulse response functions of terrestrial carbon cycle models: method and application." *Global Change Biology* 5.4, pp. 371–394 (cit. on pp. 11, 40, 48, 125).

Thornton, P. E. et al. (2009). "Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: results from an atmosphere-ocean general circulation model." *Biogeosciences* 6.10, pp. 2099–2120 (cit. on p. 102).

Tikhonov, A. N. (1963). "Solution of incorrectly formulated problems and the regularization method." In: *Dokl. Akad. Nauk*. Vol. 151, pp. 1035–1038 (cit. on p. 21).

Uppenbrink, J. (1996). "Arrhenius and global warming." *Science* 272.5265, pp. 1122–1122 (cit. on p. 1).

Van Zalinge, B. C., Q. Feng, M. Aengenheyster, and H. A. Dijkstra (2017). "On determining the point of no return in climate change." *Earth System Dynamics* 8.3, pp. 707–717 (cit. on pp. 11, 40).

Wigley, T. M. (2005). "The climate change commitment." *Science* 307.5716, pp. 1766–1769 (cit. on pp. 10, 114, 124).

Williams, R. G., A. Katavouta, and P. Goodwin (2019). "Carbon-cycle feedbacks operating in the climate system." *Current Climate Change Reports* 5.4, pp. 282–295 (cit. on pp. 4, 75, 114, 124).

Williamson, M. S., P. M. Cox, and F. J. Nijsse (2019). "Theoretical foundations of emergent constraints: relationships between climate sensitivity and global temperature variability in conceptual models." *Dynamics and Statistics of the Climate System* 3.1, dzy006 (cit. on pp. 10, 125).

Zaehle, S. et al. (2010). "Carbon and nitrogen cycle dynamics in the O-CN land surface model: 2. Role of the nitrogen cycle in the historical terrestrial carbon balance." *Global Biogeochemical Cycles* 24.1 (cit. on p. 101).

Zickfeld, K., M. Eby, H. D. Matthews, A. Schmittner, and A. J. Weaver (2011). "Nonlinearity of carbon cycle feedbacks." *Journal of Climate* 24.16, pp. 4255–4275 (cit. on pp. 4, 75).

Zwanzig, R. (2001). *Nonequilibrium statistical mechanics*. Oxford University Press (cit. on p. 5).

ACKNOWLEDGMENTS

I could only write this dissertation because I received tremendous support from many people.

First, I am greatly indebted to my supervisor Christian Reick for his wisdom and patience in showing me the ways of science. Thank you for your excellent guidance and for our deep discussions that allowed me so much fun that I could hardly believe I was working. Many thanks also to my co-supervisor Julia Pongratz for introducing me to the world of the carbon cycle by allowing me into the regular meetings of the FOM group. I had much fun during this time and learned a lot from the applied view of research of the group. Thank you also Martin Claussen for being the Panel Chair and providing excellent guidance during our Panel Meetings.

A big thank you goes to all people that have helped me by reviewing parts of this dissertation: Matteo, Sirisha, Isabella, Pinhsin, and Zoé. Thank you also to my colleagues and friends from the IMPRS and the institute for making this such a nice environment to work: Johannes, Rafaela, Sylvia, Alex, Hao-wei, Lena, Julia, Kim, Josephine, Meike, Nora, István, Arjun, Diego, David, Vimal, Katherine, Sally, and Geet. Thank you my friends from the German course who have made life in Hamburg more fun: Matheusz, Charly, and Zhe. Thanks also to all nice people from the Land department, for making this workplace a very pleasant one. Thank you Veronika and Thomas for all the times you have helped me with programming issues.

Thank you very much people from the IMPRS office: Antje, Connie, and Michaela. My life in Hamburg would have never been so easy without your invaluable support. A special thanks goes to Antje, for whom I have developed great respect and admiration for her ability of helping others in an extremely professional and human way. Thank you Antje for having helped us even outside working hours.

Meu último obrigado escrevo em português, já que é direcionado àqueles mais próximos das minhas raízes e do meu coração. Muito obrigado a todos os meus amigos do Brasil, por mesmo de longe estarem de alguma maneira por perto: Felipe, Felipe Sérgio, Pablo, Isabella, Brum, Rodrigo, Magno, Vinícius. Agradeço também ao Professor Alberto Paiva, por ter me apoiado desde o início da minha carreira acadêmica. Meu muito obrigado a minha família: mãe, pai, por terem me dado suporte em todos os momentos da minha vida, e mesmo de longe terem me apoiado todo o tempo. Stéphanie, muito obrigado também pelo seu suporte em todos os momentos difíceis. Especialmente, obrigado por ter sido tão forte nesse ano tão complicado. Gostaria de ter estado ao lado de vocês. Por fim, agradeço a Isabella, por ter topado o desafio de vir para Hamburgo comigo, e por toda a sua paciência e esforço em tornar nossa vida mais fácil aqui. Definitivamente, não teria conseguido sem você.

I thank also André Miede and Ivo Pletikosić , the authors of the CLASSICTHESIS template, which was used to typeset this document (available at: https://bitbucket.org/amiede/classicthesis/).

AFFIRMATION ON OATH
Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation mit dem Titel: "Response Characteristics of the Global Carbon Cycle in Earth System Models" selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel – insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen – benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Dissertation oder Teile davon vorher weder im In- noch im Ausland in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Hamburg, 12 August 2020

Guilherme Luiz Torres
Mendonça

Hinweis / Reference

Die gesamten Veröffentlichungen in der Publikationsreihe des MPI-M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

sind über die Internetseiten des Max-Planck-Instituts für Meteorologie erhältlich:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html

All the publications in the series of the MPI -M
„Berichte zur Erdsystemforschung / Reports on Earth System Science“,
ISSN 1614-1199

are available on the website of the Max Planck Institute for Meteorology:
http://www.mpimet.mpg.de/wissenschaft/publikationen.html