

DIMEAS  
POLYTECHNIC OF TURIN  
MASTER'S DEGREE IN AEROSPACE ENGINEERING



MASTER'S DEGREE THESIS

---

An Application of the Exact  
Response Theory to Climate System  
Perturbations

---

AUTHOR: Sonsoles Moreno Toledo  
SUPERVISOR: Lamberto Rondoni  
CO-SUPERVISOR: Carlos Mejía Monasterio

September 2022



# Contents

<b>Abstract</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Climate . . . . .	1
1.1.1 Climate Change . . . . .	3
1.1.2 Governing Equations of the Climate System . . . . .	4
1.1.3 Climate Modeling . . . . .	6
1.1.4 Climate Sensitivity and Equilibrium Climate Sensitivity (ECS) . . . . .	7
1.1.5 General Framework for Climate Response . . . . .	9
1.2 Dynamical Systems and Chaos . . . . .	10
1.2.1 Definitions . . . . .	11
1.2.2 Chaos . . . . .	12
<b>2 Ruelle’s Linear Response Theory</b>	<b>17</b>
2.1 Summary of the Theory . . . . .	17
2.2 Lucarini’s Approach to Climate Change Prediction . . . . .	19
2.3 Limitations of the Theory . . . . .	21
<b>3 Exact Response theory</b>	<b>23</b>
3.1 Introduction . . . . .	23
3.2 Mathematical Framework . . . . .	24
3.2.1 Conditions on the Dynamics and Probability Density Function . . . . .	28
3.3 Current Applications of Exact Response Theory . . . . .	29
3.3.1 Dynamics of Kuramoto Oscillators . . . . .	29
<b>4 Dynamics of the System</b>	<b>35</b>
4.1 Definition of the System . . . . .	35
4.2 Dynamics of the Resulting System . . . . .	39
4.3 Fixed Points . . . . .	40
4.3.1 Fixed Point $x_I$ . . . . .	41
4.3.2 Fixed Point $x_{II}$ . . . . .	43
4.3.3 Fixed Point $x_{III}$ . . . . .	44
4.4 Attractor . . . . .	45
4.5 Phase Space . . . . .	46

<b>5</b>	<b>Probability Density Function</b>	<b>49</b>
5.1	Uniform Distribution . . . . .	49
5.1.1	Domain Preserving its Volume . . . . .	50
5.1.2	Pseudo-arbitrary Domain . . . . .	51
5.2	3D Gaussian Distribution . . . . .	52
<b>6</b>	<b>Numerical Computation of the Exact Response Theory</b>	<b>53</b>
6.1	Perturbed Vector Field . . . . .	53
6.2	Computing the Evolution of the Observable $X$ . . . . .	54
6.2.1	Code Implementation . . . . .	55
6.2.2	Results . . . . .	55
<b>7</b>	<b>Conclusions</b>	<b>57</b>
<b>A</b>	<b>Julia code</b>	<b>59</b>

# List of Figures

1.1	Lorenz strange attractor for $\sigma = 10$ , $r = 28$ , $b = 8/3$ . . . . .	14
1.2	Evolution of $x_n$ with the generations for $r = 3.8$ . . . . .	15
3.1	Evolution of $\langle \Omega_V^{f_0} \rangle_{f_t}$ and $\langle (\Omega_V^{f_0} \circ \mathcal{S}^s) \Omega_V^{f_0} \rangle_{f_0}$ with the time when $N = 2$ , $K = 1$ and $\omega = 0$ . The averages were taken over a set of 5041 trajectories with initial data sampled from the uniform distribution $[0, 2\pi)$ . . . . .	32
4.1	Solution of the system as showed in Saltzman and Maasch (1988). . . . .	39
4.2	Evolution through time for the three variables. . . . .	40
4.3	Set of initial points around first fixed point $x_I$ . . . . .	42
4.4	Trajectories of points around the unstable fixed point $x_I$ . . . . .	42
4.5	Zoom of trajectories around unstable fixed point $x_I$ . . . . .	43
4.6	Trajectories of points around the unstable fixed point $x_{II}$ . . . . .	43
4.7	Zoom of trajectories around unstable fixed point $x_{II}$ . . . . .	44
4.8	Trajectories of points around the stable fixed point $x_{III}$ . . . . .	45
4.9	Example of trajectories being dragged to the attractor. . . . .	45
4.10	Attractor found for the dynamical system presented by Saltzman and Maasch. . . . .	46
4.11	Phase Portrait. . . . .	47
4.12	Phase Portrait. . . . .	47
4.13	Phase Portrait inside the attractor. . . . .	48
6.1	Evolution of the ensemble average of the observable $X$ after the perturbation in the parameter $r$ . . . . .	55



# Abstract

The anthropogenic contribution to the evolution of climate is currently a central topic in scientific studies. The Intergovernmental Panel on Climate Change (IPCC) assessment reports are based on the best science at hand, and summarize the progress in observing, modeling, understanding and predicting the evolution of the climate system. The climate system is forced, dissipative, nonlinear, chaotic, and out of thermodynamic equilibrium. Along with the difficulties that these characteristics come with, the climate system presents some particularities such as, the presence of well-defined subsystems, the continuous variation of the forcings, lack of scale separation, etc., which add more obstacles to the task. In this context, several theories and models are developed to predict the response of the system to perturbations. In particular, Valerio Lucarini's approach is reviewed. He makes use of Ruelle's linear response theory, which was found to be too restricted since it is limited to linear perturbations applied to Axiom A systems.

The Exact Response Theory proposed here, would overcome these limitations. It was developed within the field of Nonequilibrium Molecular Dynamics, and it is expected to predict the response of a system with many degrees of freedom even in the presence of arbitrarily large perturbations and modifications of states, allowing the study of the relaxation of particle systems to equilibrium or non-equilibrium steady states. The theory introduces the Dissipation Function,  $\Omega$ , as its basis. The Dissipation Function determines non-equilibrium properties the same way the thermodynamic potentials determine equilibrium state properties.

In order to apply the theory, it was studied the dynamical model describing the Pleistocene ice ages developed by Saltzman. That approach to the climate system reproduces the evolution of global ice mass, atmospheric  $CO_2$  concentration and North Atlantic deep water amount. There were found three fixed points and an attractor, the later responsible of the oscillations that the evolution exhibits along the time. The evolution of an observable when the system is perturbed, was computed by means of the presented theory.



# Chapter 1

## Introduction

This first introductory chapter aims to provide the appropriate background for the matters studied in the present work. Therefore, since the work developed here deals with applying the exact response theory to the climate system response, an introduction to the climate system response, the problematic of climate change and the state of the art of the developed methods to address the problem is presented. Furthermore, given that the new current of concepts and methods that are being applied to climate response and that are in particular reviewed in the present work, consists in applying the dynamical systems approach, the key concepts and definitions of this subject are exposed. Finally, there is a section dedicated to chaos, since the climate system presents chaotic behavior, as it will be showed later.

### 1.1 Climate

The climate system is forced, dissipative, non-linear, chaotic and out of thermodynamic equilibrium. The interplay of positive and negative feedbacks, instabilities and saturation mechanisms are the reason of its complex natural variability. These processes span a broad range of temporal and spatial scales and combine many chemical species as well as all of the most common physical phases. The heterogeneous phenomenology of the climate system comprises the microphysics of clouds, cloud-radiation interactions, atmospheric and oceanic boundary layers, and several scales of turbulence. Moreover, it evolves under the action of large-scale agents that drive and modulate its evolution, mainly differential solar heating and the Earth's rotation and gravitation. In addition to that, the complexity of the physics is mixed with the chaotic character of the dynamics. Furthermore, the large natural variability of the system is greatly affected by relatively small changes in the forcing, which could be anthropogenic or natural.

The climate system's governing equations are presented in subsection 1.1.2. At a macroscopic level, climate is driven by differences in the absorption of solar radiation throughout the depth of the atmosphere, and in a narrow surface layer of both the ocean and the soil. The large-scale circulation of the atmosphere is to first or-

der originated by horizontal and vertical fluxes resulting from the gradients in solar radiation absorption. While the ocean circulation is set by surface or near-surface transfers of mass, momentum and energy with the atmosphere. At steady state, the convergence of enthalpy transported by both the atmosphere and the ocean compensates for the radiative imbalance at the top of the atmosphere.

In Lorenz (1967), it was developed the classical theory of the general atmosphere's circulation. The mechanisms of energy generation, conversion and dissipation that produce the circulation are explained in detail. According to Lorenz, the large-scale flows are originated by the conversion of available potential energy, resulting of the differential heating of the atmosphere, into kinetic energy. The energy cycle would be completed by energy cascading to smaller scales to eventually be dissipated.

In general, the climate system transforms radiative heat into mechanical energy. The conversion happens through fairly three-dimensional baroclinic instabilities, that are generated by large temperature gradients. They give rise to a negative feedback, since they tend to decrease these temperature gradients by favoring the mixing between masses and fluids at different temperatures.

As stated in Ghil and Lucarini (2019), *“the closure of the coupled thermodynamical equations governing the general circulation of the atmosphere and ocean would provide a self-consistent theory of climate. Such a theory should be able to connect instabilities and large-scale stabilizing processes on longer spatial and temporal scales, and to predict its response to a variety of forcings, both natural and anthropogenic. This goal is being actively pursued but is still out of reach at this time”*.

Analyzing the climate system's entropy budget gives one a good global perspective of the system. The Earth as a whole absorbs shortwave radiation carried by low-entropy solar photons, and emits to space infrared radiation via high-entropy thermal photons. Apart from the viscous dissipation of kinetic energy, a lot more of irreversible processes contribute to the total material entropy production. Some of these processes are turbulent diffusion of heat and chemical species, irreversible phase transitions associated with hydrological processes, chemical reactions involved in the biogeochemistry of the planet. The production of entropy is balanced by a net outgoing flux of entropy leaving mainly through the top of the atmosphere.

The phenomenology of the climate system is normally approached by focusing in different and complementary aspects, being included the following:

- Wavelike features, like Rossby waves or equatorially trapped waves, playing a fundamental role in the transport of energy, momentum, and water vapor, and also in the study of oceanic, atmospheric and coupled-system predictability.
- Particlelike features, like hurricanes, extratropical cyclones and oceanic vor-

tices, that strongly affect the local properties of the system.

- Turbulent cascades, which are crucial in the development of large eddies, as well as in the mixing and dissipation within the planetary boundary layer.

Each of these perspectives is useful, but also, they overlap and complement each other. However, none of them is able to provide by itself a comprehensive understanding of the properties of the climate system.

In addition to the inherent difficulties to understand and predict the dynamics of nonlinear, complex systems out of equilibrium, the climate system presents the following additional obstacles:

- The existence of well-defined subsystems (atmosphere, ocean, cryosphere) with different physical and chemical properties and widely diverse timescales and space scales. Along with the complex processes coupling these subsystems.
- The continuous alteration of the atmospheric composition, caused by the varying set of forces resulting from fluctuations in the incoming solar radiation and in the processes, natural and anthropogenic.
- The lack of scale separation between the different processes.
- The lack of climatic fields observations that are detailed, homogeneous, with high-resolution and lasted for a long time.
- The fact that only one realization of the processes that give rise to the climate evolution is available.

Consequently, it is hard to separate the climate system's response to different forcings from its natural variability. This difficulty is a considerable handicap towards an unified theory of climate evolution, however some new promising ideas are emerging to overcome it and they are treated in subsection 1.1.5. As well as in section 2.2, with the approach of Valerio Lucarini to the problem.

### 1.1.1 Climate Change

The interest in climate research is not merely scientific, in fact, recent interest comes from the accumulated observational and modeling evidence of the anthropogenic influence in the climate system. It has become a hot topic since its consequences are already affecting the whole population of the world, even if the degree of the impact is different depending on the pertinent place in the world and the social class, what was denoted as climate injustice. However, it is a very polemic topic since mitigating climate change and acknowledging its causes, questions the economic system and the first-world lifestyle, and therefore, it is easy to find people calling scientific data into question and not taking the predicted and observed consequences seriously,

politicians and media included. The lack of meaningful actions and progress made by the governments has led to the growth of young-people-driven movements like Extinction Rebellion and Fridays For Future.

The United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO) established in 1988 the Intergovernmental Panel on Climate Change (IPCC) in order to review and coordinate the research activities in this respect. They issue the assessments reports (ARs) every 4-6 years, where they summarize the scientific progress, the open questions, and the bottlenecks regarding the capacity to observe, model, understand and predict the evolution of the climate system. The IPCC Working Group I is the one focused on the physical basis of climate change, the topic addressed in the present work. The IPCC assessment reports are based on the best available science, being policy relevant but not policy prescriptive. They are supposed to guarantee neutrality although they must pass the check of the IPCC's 195 member countries before going public. Hence, their release leads to considerable and often adversarial debates involving diverse stakeholders from science, politics, civil society and business.

### 1.1.2 Governing Equations of the Climate System

In order to describe the evolution of the climate subdomains, i.e. the atmosphere, the ocean, the soil, and the ice masses, the continuum approximation can be used. That way, each subsystem is represented by field variables that depend on three spatial dimensions and time. For each subdomain, the following field variables are considered: the density  $\rho$ , the heat capacity at constant volume  $c$ , the concentration of the chemical species ( $\xi_k$ ,  $k = 1, \dots, K$ ) contained in the medium and present in different phases, the three components of the velocity vector ( $v_i$ ,  $i = 1, 2, 3$ ), the temperature  $T$ , the pressure  $p$ , the heating rate  $J$ , and the gravitational potential  $\Phi$ .

Considering the thin-shell approximation, the gravitational potential at the local sea level can be assumed to be null, and hence it is approximated as  $\Phi = gz$ , called geopotential now, where  $g$  is the gravity at the surface and  $z$  is the geometric height above the sea level. Furthermore, as the climate system is embedded in a non-inertial frame of reference, its angular velocity has to be taken into account,  $\Omega$  with components ( $\Omega_i$ ,  $i = 1, 2, 3$ ).

The Partial Differential Equations (PDEs) that govern the evolution of the field variables are based on the mass, momentum and energy budgets. If the fluid contains different chemical species, their separate budgets must be accounted for as well.

The mass budget for the contained species is described by the following equation:

$$\partial_t(\rho\xi_k) = -\partial_i(\rho\xi_k v_i) + \mathbf{D}_{\xi_k} + \mathbf{L}_{\xi_k} + \mathbf{S}_{\xi_k}, \quad (1.1)$$

where  $\partial_t$  is the partial derivative in time,  $\partial_i$  is the partial derivative in the  $x_i$  direction, and  $\mathbf{D}_{\xi_k}$ ,  $\mathbf{L}_{\xi_k}$ , and  $\mathbf{S}_{\xi_k}$  are respectively the diffusion operator, phase changes, and local mass budget as a result of other chemical reactions associated to  $k$ .

The  $i$ th component of the momentum budget is given by:

$$\partial_t(\rho v_i) = -\partial_j(\rho v_j v_i) - \partial_i p + \rho \partial_i \Phi - 2\rho \epsilon_{ijk} \Omega_j v_k + T_i + F_i, \quad (1.2)$$

where  $\epsilon_{ijk}$ , the Levi-Civita antisymmetric tensor, is used to write the Coriolis force,  $T_i$  represents direct mechanical forcings,  $F_i = -\partial_j \tau_{ij}$  is the friction ( $\tau_{ij}$  is the stress tensor).

For the atmosphere and the ocean, the two fluid envelopes of Earth, the following general state equation can be used

$$\rho = g(T, p, \xi_1, \dots, \xi_K). \quad (1.3)$$

In general, as a first approximation,  $K = 1$  may be considered, thus  $\xi$  would be moisture in the atmosphere and salinity in the ocean. Neglecting reactions other than the phase changes between liquid and gas phases of water, the equation of the specific energy in the atmosphere is

$$e = c_v T + \Phi + v_j v_j / 2 + Lq. \quad (1.4)$$

Here,  $c_v$  is the specific heat at constant volume for the gaseous mixture,  $L$  is the latent heat of evaporation, and  $q = \rho \xi$  is the specific humidity. For the ocean the obtained equation is

$$e = c_w T + \Phi + v_j v_j / 2, \quad (1.5)$$

where  $c_w$  is the specific heat at constant volume of the water, neglecting the effects of salinity and pressure. Lastly, for the soil it is used,  $e = c_s T + \Phi$ , and for the ice similarly,  $e = c_I T + \Phi$ .

Finally, the following general equation for the local energy budget is derived:

$$\partial_t(\rho e) = -\partial_j(\rho \varepsilon v_j) - \partial_j Q_j^{SW} - \partial_j Q_j^{LW} - \partial_j J_j^{SH} - \partial_j J_j^{LH} - \partial_j(v_i \tau_{ij}) + v_i T_i, \quad (1.6)$$

where  $e$  is the energy per unit mass and  $\varepsilon = e + p/\rho$  is the enthalpy per unit mass. The energy sinks and sources may be written as the sum of the work done by the mechanical forcing ( $v_i T_i$ ) and of the respective divergences of the shortwave (solar) and longwave (terrestrial) components of the Poynting vector  $Q_j^{SW}$  and  $Q_j^{LW}$ , of the turbulent sensible and latent heat fluxes  $J_j^{SH}$  and  $J_j^{LH}$ , and the scalar product of the velocity field with the stress tensor ( $v_i \tau_{ij}$ ).

The presence of non-homogeneous absorption of shortwave radiation given the geometry of the Sun-Earth system and the physicochemical properties of the climatic subdomains induces the presence of non-equilibrium conditions for the climate system.

### 1.1.3 Climate Modeling

A main area of interest in the climate sciences, and also the topic addressed in the present work, is the developing of numerical methods for simulating the past, present and future of the climate system. There are huge differences between the climate models in terms of scientific scope, computational cost, and flexibility. Thus, a hierarchy of climate models is to be considered, rather than one model that would include all subsystems, processes and scales.

Models of distinct levels of complexity and detail are suited for addressing different kinds of inquiries, depending on the main spatial and temporal scales of interest. At the top of the hierarchy, the global climate models (GCM) are found. These model's goal is to represent, at the highest computationally attainable resolution, the biggest number of physical, chemical and biological processes of the Earth system. The atmosphere and the ocean are at the core of the Earth system models being developed today. Their modeling relies on the governing equations presented before in subsection 1.1.2. Later in the section models at the bottom of the hierarchy are presented.

Climate modeling and prediction deal with several kinds of uncertainties, beginning with the uncertainties in predicting the state of the system at a certain lead time, due to the uncertain knowledge of the present state. Geophysical flows are commonly chaotic, as well as other processes in the system. Therefore, the climate system depends sensitively on its initial data, as acknowledged by Lorenz. Chaotic dynamical systems are explained in more in detail in subsection 1.2.2.

A second kind of uncertainties affects the determination of the system's statistical properties, that is, its mean state and variability, and its response to forcings of a different nature. These uncertainties correspond to the uncertainties in model formulation, and the unavoidably restricted knowledge of the external forcings. They affect critically the modeling of abrupt changes in the climate and the processes that could lead to them.

Finally, from trying to find the best metrics for analyzing a model's outputs and evaluating its skills, comes the third kind of uncertainties. As there is no a priori valid criterion for choosing a good climatic observable, and so there is no unique method for testing climate models.

#### **Energy Balance Methods (EBM)**

Currently, the best developed hierarchy is for atmospheric models. At the first rung of this hierarchy, one finds the zero-dimensional (0D) models. Where the number of dimensions (from 0 to 3), refers to the number of independent space variables that are used to describe the domain of the model (physical-space dimensions). These models receive the name of Energy Balance Models (EBM). The 0D models basically try to follow the evolution of the globally averaged surface air temperature, resulting from changes in the global radiative balance. That way the following expression are

considered:

$$c \frac{d\bar{T}}{dt} = R_i - R_o, \quad (1.7)$$

$$R_i = \mu Q_0 (1 - \alpha(\bar{T})), \quad (1.8)$$

$$R_o = \sigma m(\bar{T})(\bar{T})^4, \quad (1.9)$$

where  $R_i$  is the incoming solar radiation,  $R_o$  is the outgoing terrestrial radiation. The heat capacity  $c$  is that of the global atmosphere plus some fraction (or whole) of the global ocean capacity, depending on the timescale of interest.  $\frac{d\bar{T}}{dt}$  gives the rate of change of  $\bar{T}$  with time.  $Q_0$  is the solar radiation received at the top of the atmosphere (solar constant),  $\sigma$  is the Stefan-Boltzmann constant,  $\mu$  is an insolation parameter,  $\alpha$  is the albedo and  $m$  is the grayness factor.

For 1D atmospheric methods, there are two kinds, depending on the single spatial variable which could be the latitude or height. They take Equation 1.7 and generalize it for a evolution of the temperature like  $T = T(x, t)$ , thus

$$c(x) \frac{\partial T}{\partial t} = R_i - R_o + D. \quad (1.10)$$

Here, the terms  $R_i$  and  $R_o$  are similar to those presented in Equation 1.8 and Equation 1.9, but now they can depend on the meridional coordinate  $x$  (latitude, colatitude or sine of latitude), along with time  $t$  and temperature  $T$ . The horizontal heat-flux term  $D$  describes the convergence of the heat transport across latitude belts (typically containing first and second partial derivatives of  $T$  with respect to  $x$ ), whereas  $c(x)$  is the system's space-dependent heat capacity. In this model, the rate of change of local temperature with respect to time also becomes a partial derivative,  $\frac{\partial T(x,t)}{\partial t}$ , since Equation 1.10 corresponds to a nonlinear heat or reaction-diffusion equation from the physical point of view, and to a nonlinear parabolic PDE from the mathematical one.

Already using the 0D EBM several stable steady states are found, finding the bistability of the system. While the complexity of the models grow, so does its behavior. Apart from multiple equilibria, complex processes give rise to the system's internal variability, by successive instabilities setting in, competing and eventually leading to the chaotic nature of the climate's evolution.

#### 1.1.4 Climate Sensitivity and Equilibrium Climate Sensitivity (ECS)

The goal of climate sensitivity is to measure the response of the climate system to external perturbations of Earth's radiative balance. This measure is used to obtain the

changes in mean temperature over the century as a response of increasing concentrations of atmospheric greenhouse gases. In order to exemplify, it is considered the 0D EBM from subsection 1.1.3, where the net radiation at the top of the atmosphere depends on the average temperature near the Earth's surface as  $R = R(T)$ . It considers longwave and shortwave processes so that  $\frac{cdT}{dt} = R(T)$ . In addition to that, it is assumed that there are  $N$  climatic variables,  $(\alpha_k = \alpha_k(T), k = 1, \dots, N)$  that at first approximation are directly affected only by the temperature change and that are able to affect the radiative balance. Thus, it can be written  $R = R(T, \alpha_1(T), \dots, \alpha_N(T))$ . Moreover, it is assumed that for a certain reference temperature  $T_0$ , the net radiation is null,  $R(T_0) = 0$ , corresponding to steady-state conditions. Now, the simplest framework for climate sensitivity is thinking of the difference in global annual mean surface air temperature  $\Delta T$  between two statistical steady states presenting different  $CO_2$  concentration levels. Then, it is assumed that the change in  $CO_2$  concentration is translated into applying an extra radiative forcing  $\Delta\tilde{R}$  to the system. And the corresponding  $\Delta T$  is searched, so that  $R(T_0 + \Delta T) + \Delta\tilde{R} = 0$ .

For small  $\Delta T$  and smooth  $R = R(T)$ , introducing the following notations:

$$\frac{1}{\lambda_0(T_0)} = -\left.\frac{\partial R}{\partial T}\right|_{T=T_0}, \quad (1.11)$$

$$f_k(T_0) = -\lambda_0(T_0)\left.\frac{\partial R}{\partial \alpha_k}\frac{\partial \alpha_k}{\partial T}\right|_{T=T_0}, \quad (1.12)$$

and considering the Taylor expansion, it yields for the ‘‘reference sensitivity’’  $\lambda_0$  and the ‘‘feedback factors’’  $f_k$  at the reference state  $T = T_0$

$$\Delta\tilde{R} = \frac{1 - \sum_{k=1}^N f_k(T_0)}{\lambda_0(T_0)}\Delta T + \mathcal{O}((\Delta T)^2) \quad (1.13)$$

One of the several measures of climate sensitivity is the so-called equilibrium climate sensitivity (ECS), which denotes the globally and annually averaged surface air temperature increase that would result from continuously doubling of the  $CO_2$  concentration in Earth's atmosphere versus the one of the reference state after the climate system reaches a new steady-state equilibrium. Resulting,

$$ECS = \frac{\lambda_0(T_0)}{1 - \sum_{k=1}^N f_k(T_0)}\Delta\tilde{R}_{2\times CO_2} \quad (1.14)$$

The climate sensitivity concept can be generalized to describe linear dependence of the long-term average of any climatic observable with respect to the radiative forcing resulting from changes in  $CO_2$  or other greenhouse gases, along with changes in solar radiation, aerosol concentration, or any other change in the forcings.

The ECS is universally considered to be the most important indicator, in particular it is a state-dependent indicator, in understanding the climate response to forcings both natural and anthropogenic. However, even if it is useful, the ECS concept finds some practical difficulties since the definition assumes that after the forcing is applied, the climate reaches a new steady state after all transients have died

out. And, given the multiscale for both the time and space that the climate system has, it is far from trivial to define an effective cutoff timescale that would include all transient behavior. Hence, it is needed to associate each ECS estimate to a reference timescale.

### Transient Climate Response (TCR)

Transient climate response (TCR) is defined as the change in the globally averaged surface air temperature recorded at the time at which  $CO_2$  has doubled due to an increase at 1% annual rate, that is approximately after 70 years, having started at a given reference value  $T_0$ . It is closer to capture the evolution in time of climate change, as it addresses the transient rather than the asymptotic response of the system. Being that the operational definition agrees well with the standard IPCC-like simulation protocols, the TCR is better to test model outputs against observational datasets from the industrial era, than ECS.

### 1.1.5 General Framework for Climate Response

The ECS comes with some limitations, such as addressing only long-term changes, no spatial information, the impossibility to difference between radiative forcings from distinct physical and chemical processes. Therefore, some new concepts and methods are recently added to climate science, from the field of non-autonomous and random dynamical systems. As explained in Ghil and Lucarini (2019), *“The setting of non-autonomous and of stochastically forced dynamical systems allows one to examine the interaction of internal climate variability with the forcing, whether natural or anthropogenic; it also helps provide a general definition of climate response that takes into account the climate system’s non-equilibrium behavior, its time-dependent forcing, and its spatial patterns.”* The concept of Pullback Attractors arises in these new methods.

### Pullback Attractors (PBAs)

Considering a continuous-time dynamical system

$$\dot{x} = F(x, t), \tag{1.15}$$

on a compact manifold  $\mathcal{Y} \subset \mathbb{R}^d$ , where  $x(t) = \phi(t, t_0)x(t_0)$ , with  $x(t = t_0) = x_{in} \in \mathcal{Y}$  initial condition and  $\phi(t, t_0)$  is defined for all  $t \geq t_0$  with  $\phi(s, s) = \mathbf{1}$ . The interest is in forced and dissipative systems that, with probability 1, initial states in the distant past are attracted at time  $t$  toward  $\mathbf{A}(t)$ . Being  $\mathbf{A}(t)$  a time-dependent family of geometrical sets that define the system’s pullback attractor. In formal terms, a family of objects  $\cup_{t \in \mathbb{R}} A(t)$  in the finite-dimensional, complete metric phase space  $\mathcal{Y}$  is a pullback attractor for the system given by Equation 1.15 if the following conditions are obeyed:

- $\forall t$ ,  $A(t)$  is a compact subset of  $\mathcal{Y}$  which is covariant with the dynamics, i.e.  $\phi(s, t)A(t) = A(s)$ ,  $s \geq t$ .

- $\forall t \lim_{t_0 \rightarrow -\infty} d_{\mathcal{Y}}(\phi(t, t_0)B, A(t)) = 0$  for a.e. measurable set  $B \subset \mathcal{Y}$ .

Where  $d_{\mathcal{Y}}(P, Q)$  is the Hausdorff semi-distance between the  $P \subset \mathcal{Y}$  and  $Q \subset \mathcal{Y}$ .

The PBA may also be constructed when random forcing is present. An example of numerical application of PBAs was performed in several studies to explain the wind-driven circulation and thermohaline circulation.

As in the most recent IPCC reports and according to the standard protocols, future climate projections are virtually always performed taking as initial states the final states of sufficiently long simulations of historical climate conditions, it is acceptable to assume that the pullback time  $\tau$  is large enough and the covariance properties of the associated  $\mathbf{A}(t)$  sets are therefore well approximated. Being the pullback time defined as the time that a set  $\mathbf{B}$  initialized at  $t = t_0 - \tau$ , takes to evolve to  $\mathbf{A}(t_0)$ .

### Fluctuation Dissipation Theorem (FDT)

The fluctuation dissipation theorem was born in the classical theory of many-particle systems in thermodynamics equilibrium. Basically it states that the return to equilibrium of the system will be the same both ways if the perturbation that modified the system's state is due to a small external force or to an internal, random fluctuation. The FDT has been applied by some scientists to the output of climate models in order to predict the response to a step-like increase of the solar irradiance, increases in atmospheric  $CO_2$  concentration and to predict the response of an atmospheric model to localized heating anomalies. It was also used to reduced the uncertainty in ECS. An important issue of these studies, is the limitation given by the fact that they take a Gaussian approximation for the invariant measure, which implies that the climate would be in thermodynamics equilibrium, which addresses the applicability of the FDT in a reduced phase space.

## 1.2 Dynamical Systems and Chaos

Dynamics is the subject that deals with systems which evolve in time. The subject was born, as explained in Strogatz (2018), “*when Newton invented differential equations, discovered his laws of motion and universal gravitation, and combined them to explain Kepler’s laws of planetary motion*”, in the mid-17th century. More in detail, Newton solved the two-body problem (motion of the Earth around the Sun, for example, given the gravitational attraction law between them). However, when the three-body was later studied, it was concluded that there was no way of solving it analytically. It was not until the late 1800s when thanks to Poincaré’s work, there was a step-forward. He changed the point of view to a more qualitative point of view, questioning for example “*Is the solar system stable forever, or will some planets eventually fly off to infinity?*”, instead of trying to determine the exact positions of the planets at all times. Besides, he was the first person in having a brief insight into chaos.

With the arrival of the high-speed computer in the 1950s, there were big advances in the subject of dynamics. Now, the equations could be experimented in a revolutionary way, and some intuition about nonlinear systems was developed. Such experiments led Lorenz to discover in 1963 the chaotic motion on a strange attractor. The particularities of chaos will be introduced in subsection 1.2.2.

### 1.2.1 Definitions

In the next lines, some basic definitions related to dynamics and relevant to the dynamical system treated in this work, will be presented, as found in Strogatz (2018). To begin with, in the work developed here, the dynamical systems that are treated are differential equations. Differential equations describe the evolution of systems in continuous time. There exist other type, the iterated maps, used in problems where time is discrete.

If a general system is considered, i.e.  $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ , being  $\mathbf{x} = (x_1, \dots, x_N)$ , an abstract space constructed with these coordinates  $(x_1, \dots, x_N)$ , is called the **phase space**. An arbitrary initial condition  $\mathbf{x}_0$ , which is called **phase point**, will evolve following a certain function  $\mathbf{x}(t)$ . This function is called **trajectory** based at  $\mathbf{x}_0$  and it reproduces the solution of the differential equations beginning from the initial condition  $\mathbf{x}_0$ . A picture showing all the possible trajectories of the system in a qualitative way is called a **phase portrait**.

In dynamics the differential equations are interpreted as a vector field. Thus, in a general first-order system, i.e.  $\dot{x} = f(x)$ , the differential equation (vector field) dictates the velocity vector  $\dot{x}$  at each  $x$ .

#### Fixed Points

When the condition  $\dot{\mathbf{x}} = \mathbf{0}$  is fulfilled, the fixed points are found. If the vector field is imagined as a flow, these points are the stagnation points of the flow. The fixed points represent equilibrium solutions in the differential equation. There are two different types of fixed points, the stable ones and the unstable ones. The stable fixed points are also called attractors or sinks, since the flow is directed towards them. Instead, the unstable fixed points are called repellers or sources, as the flow is not rejected from them. An important note is that the stability of the fixed points is defined based on small disturbances, which means that the stable fixed points are *locally* stable.

If the fixed point is denoted as  $\mathbf{x}^*$ , the various types of stability are:

- **Attracting:** A fixed point is attracting if there is a  $\delta > 0$  such that  $\lim_{t \rightarrow \infty} \mathbf{x}(t) = \mathbf{x}^*$  whenever  $\|\mathbf{x}(0) - \mathbf{x}^*\| < \delta$ . Thus, all the trajectories that start within a distance of  $\delta$  from  $\mathbf{x}^*$ , it is guaranteed that will converge to  $\mathbf{x}^*$  eventually.
- **Liapunov stable:** The fixed point  $\mathbf{x}^*$  would be Liapunov stable if for each  $\epsilon > 0$ , there is a  $\delta > 0$  such that  $\|\mathbf{x}(t) - \mathbf{x}^*\| < \epsilon$  whenever  $t \geq 0$  and

$\|\mathbf{x}(0) - \mathbf{x}^*\| < \delta$ . Hence, the trajectories starting within  $\delta$  of  $\mathbf{x}^*$  remain within  $\epsilon$  of  $\mathbf{x}^*$  for all positive time.

- **Asymptotically stable:** A fixed point is asymptotically when it is both attracting and Liapunov stable.

## Limit Cycles

A limit cycle is an isolated close trajectory, meaning that near trajectories are not closed. They can spiral towards the limit cycle or away from it. If all the neighboring trajectories direct towards the limit cycle, it is said to be stable or attracting. In the other cases it will be unstable or half-stable. In particular, stable limit cycles are present in model systems that exhibit self-sustained oscillations. Limit cycles are nonlinear phenomena, as they cannot appear in linear systems.

### 1.2.2 Chaos

Lorenz studied a simplified model of convection rolls in the atmosphere, in order to achieve some insight into the well-known unpredictability of the weather. He found that the solutions to his equations never settled down to equilibrium or to a periodic state, instead they continued to oscillate in an irregular, aperiodic shape. Furthermore, if the simulation was initiated from slightly different initial conditions, the resulting behavior would start to become completely different really quick. Lorenz also exposed that there was structure in the chaos, he found the fractal were the solutions fell, in particular it was a butterfly-shaped set of points.

#### Definition

Although there is no universally accepted definition of chaos, there are three constituents generally agreed:

- “Aperiodic long term behavior”: there are trajectories which do not settle down to fixed points, periodic or quasiperiodic orbits at  $t \rightarrow \infty$ .
- “Deterministic system”: the system has no random or noisy inputs or parameters. Thus, the irregular behavior comes from the system’s nonlinearity.
- “Sensitive dependence on initial conditions”: nearby trajectories separate exponentially.

Combining all three, the following definition given by Strogatz (2018) may be written, “*Chaos is aperiodic long-term behavior in a deterministic system that exhibits sensitive dependence on initial conditions.*”

### Attractor

Roughly speaking an attractor is a set to which all nearby trajectories converge. The previous presented fixed points and limit cycles are examples of attractors. In a precise way, an attractor may be defined, following Strogatz (2018), as a closed set  $A$  with the following properties:

1.  $A$  is an invariant set: any trajectory  $\mathbf{x}(t)$  starting in  $A$  stays in  $A$  for all time.
2.  $A$  attracts an open set of initial conditions: there is an open set  $U$  containing  $A$  such that if  $\mathbf{x}(0) \in U$ , then the distance from  $\mathbf{x}(t)$  to  $A$  tends to zero as  $t \rightarrow \infty$ . This means that  $A$  attracts all trajectories that start sufficiently close to it.
3.  $A$  is minimal: there is no proper subset of  $A$  that satisfies conditions 1 and 2.

A strange attractor is an attractor that exhibits sensitive dependence on initial conditions. They are called strange attractors since normally they are fractal sets, however this geometric property does not receive such a relevance now.

### Liapunov Exponents

Related to sensitive dependence on initial conditions, arise the Liapunov Exponents. For a  $n$ -dimensional system, considering the evolution of an infinitesimal sphere of perturbed initial conditions, it is found that the given sphere becomes distorted into a infinitesimal ellipsoid. If  $\delta_k(t), k = 1, \dots, n$ , denote the length of the  $k$ th principal axis of the ellipsoid, then, one has  $\delta_k(t) \sim \delta_k(0)e^{\lambda_k t}$ , where  $\lambda_k$  are the Liapunov exponents. For large  $t$ , the diameter of the ellipsoid is controlled by the most positive  $\lambda_k$ , which is often called Liapunov exponent. When a system possesses a positive Liapunov exponent, there is a time horizon beyond which prediction breaks down, as small uncertainties are amplified really fast.

### Lorenz equations

From the aforementioned simplified model of convection rolls in the atmosphere, in Lorenz (1963) the following three-dimensional system was derived:

$$\dot{x} = \sigma(y - x) \tag{1.16}$$

$$\dot{y} = rx - y - xz \tag{1.17}$$

$$\dot{z} = xy - bz \tag{1.18}$$

where the Prandtl number  $\sigma$ , the Rayleigh number  $r$  and  $b$  are the defined positive parameters of the system. Over a wide range of parameters, the solutions of the system oscillate irregularly, although they remain in a bounded region of the phase space, they never repeat the exact trajectory. In particular, the trajectories settle in a strange attractor, which is a fractal. The fractal is a set of points with zero volume but infinite surface area. As it could be seen in Figure 1.1, the Lorenz

attractor was computed for the typical values of the parameters.

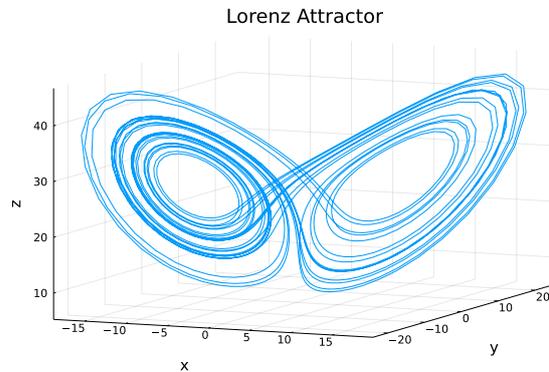


Figure 1.1: Lorenz strange attractor for  $\sigma = 10$ ,  $r = 28$ ,  $b = 8/3$ .

The Lorenz system shows nonlinearity as well as, symmetry, since replacing  $(x, y) \rightarrow (-x, -y)$  results in the same solution by the equations. Moreover, the system is dissipative, as the volumes in phase space shrink.

### Logistic Map

In May (1976), Robert May showed how simple nonlinear maps may lead to very complicated, chaotic dynamics, becoming an archetypal example of it. The logistic map was developed as a discrete-time demographic model analogous to the logistic equation (by Pierre Franois Verhulst). It is written in the following way:

$$x_{n+1} = rx_n(1 - x_n), \quad (1.19)$$

where  $x_n \geq 0$  is a dimensionless measure of the population in the  $n$ th generation and  $r \geq 0$  is the intrinsic growth rate.

Studying the behavior for different values of  $r$ , in particular in a range  $0 \leq r \leq 4$ , different types of behavior are found.

For  $r < 1$ , the population always goes extinct, as  $x_n \rightarrow 0$  when  $n \rightarrow \infty$ . While, for  $1 \leq r \leq 3$  the population grows, reaching eventually a nonzero steady state. For values of  $r$  higher than 3, the population oscillates. First, with a “period-2 cycle”, as  $x_n$  repeats every two iterations. Then, as  $r$  grows, there are found the “period-doublings”, that is, the  $x_n$  begins to repeat every 4 iterations, then every 8, etc. Until  $r$  reaches a particular value denoted  $r_\infty \approx 3.56995$  where the period corresponds to infinity. For values larger than  $r_\infty$ , the chaotic behavior is found, as shown in Figure 1.2, where the evolution of  $x_n$  along the generations was computed using a value of  $r$  bigger than the  $r_\infty$ , in particular  $r = 3.8$ .

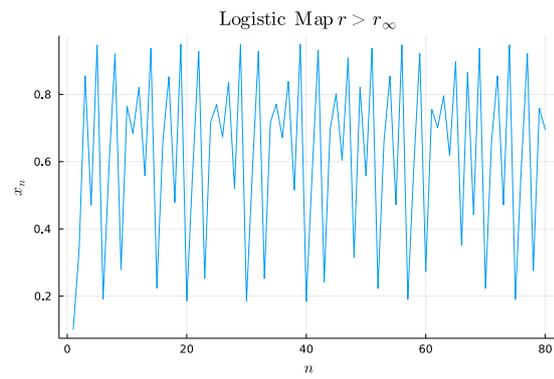


Figure 1.2: Evolution of  $x_n$  with the generations for  $r = 3.8$ .



## Chapter 2

# Ruelle's Linear Response Theory

In order to account for systems out of equilibrium, like the climate, some generalizations of the fluctuation-dissipation theorem (FDT) were developed since 1950s. In Ruelle (1998), the author David Ruelle changed the point of view, considering the problem in the setting of dynamical-systems theory, instead of statistical mechanics.

Ruelle developed a theory for the linear response of smooth uniformly hyperbolic dynamical systems (Axiom A systems). It extended the classical idea based on equilibrium time correlation functions to nonequilibrium steady states. The dissipative systems are characterized by invariant distributions which are singular with respect to the Lebesgue measure. Thus, a perturbation could cause a distribution that is singular with respect to the invariant one. Hence, when dealing with hyperbolic systems, the response is composed of two contributions, one in behalf of the expanding directions and the other due to the contracting directions. The expanding directions correspond to the unstable ones while the contracting directions to the stable ones, which contribution is responsible for relaxation to the unperturbed state.

While this theory is considered the testbed for all the other response theories, it faces some difficulties when applied to physics. Since there exist only a limited number of systems that are uniformly hyperbolic, and in addition to that, normally, the directions of stable and unstable manifolds cannot be disentangled from each other.

### 2.1 Summary of the Theory

A summary of Ruelle's linear response theory is presented in this section as in Rondoni and Dematteis (2016). If the theory is applied to systems in which the stable and unstable manifolds can be precisely identified, the following can be applied.

Since the probability measure is smooth along the unstable direction, and an impulse along that given direction results in an equally smooth distribution, for the unstable manifold, the classical Green-Kubo linear response theory can be applied.

The theory may be found in some books like Kubo et al. (1991). The Green-Kubo linear response theory yields the linear response caused by a perturbation as a time correlation function of the system at equilibrium.

As one may expect, this needs to be complemented by the contribution in behalf of the stable direction. However, this contribution is not expressed by a correlation function. Thus, some susceptibilities are obtained showing two different kinds of resonances (unstable and stable ones).

In particular, considering a map:

$$\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t)$$

with  $t \in \mathbb{Z}$ ,  $\mathbf{x} \in \mathcal{M}$ , where  $\mathcal{M}$  is the phase space and  $\mathbf{F}$  is smooth but not necessarily invertible. If the dynamics are chaotic, mixing and associated with an ergodic Sinai-Ruelle-Bowen (SRB) measure, the time average of an observable:

$$\bar{O} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} O(\mathbf{F}^t(\mathbf{x})),$$

equals the ensemble average:

$$\langle O \rangle = \int \rho_{\mathbf{F}}(d\mathbf{x}) O(\mathbf{x}) = \lim_{t \rightarrow \infty} \int d\mathbf{x} O(\mathbf{F}^t(\mathbf{x})).$$

Now, introducing a time dependent perturbation, the new dynamics are:

$$\tilde{\mathbf{x}}_{t+1} = \tilde{\mathbf{F}}_t(\tilde{\mathbf{x}}_t) + \epsilon_{t+1}(\tilde{\mathbf{x}}_t),$$

where  $\epsilon_{t+1}$  accounts for the perturbation. Since the linear response to generic perturbations can be expressed as a linear combination of impulse responses at different times, it is convenient to have in mind the impulse response:

$$\mathcal{X}_{O\epsilon}(t - \tau) = \langle O \rangle_t - \langle O \rangle = \int \rho_{\mathbf{F}}(d\mathbf{x}) \nabla(O \circ \mathbf{F}^{t-\tau})(\mathbf{F}(\mathbf{x})) \cdot \epsilon(\mathbf{x}).$$

That way, expressing the time dependent perturbation as a sum of impulses,  $\epsilon_t(\mathbf{x}) = \sum_{\tau=-\infty}^t \epsilon_{\tau}(\mathbf{x})$ , the linear response results:

$$\langle \delta O \rangle_t = \sum_{\tau=-\infty}^t \mathcal{X}_{O\epsilon}(t - \tau).$$

The series converges for uniformly hyperbolic systems, as it has been proven by Ruelle. More in detail, the unstable response converges because of exponential mixing and is given by a correlation function. Whereas the stable part, converges due to the phase space contraction.

## 2.2 Lucarini's Approach to Climate Change Prediction

In this section it is introduced the use of the Ruelle's linear response theory that Valerio Lucarini does to predict climate change, as he explains in Ghil and Lucarini (2019). Where they state that Ruelle's theory is "*an efficient and flexible tool for calculating climate response to small and moderate natural and anthropogenic forcings*". Lucarini's approach is part of the new methods incorporating the dynamical systems machinery into the climate science.

As mentioned in this article, it is a common practice to assume that a time-dependent measure  $\mu_t(d\mathbf{x})$  associated with the evolution of the dynamical system  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, t)$  exists. Yet, computing the expectation value of measurable observables with respect to this measure requires normally big efforts. They propose a "*minor generalization*" of the chaotic hypothesis found in (Gallavotti and Cohen (1995)), assuming that when considering sufficiently high-dimensional, chaotic and non-autonomous dissipative systems, the corresponding measure  $\mu_t(d\mathbf{x})$  is of SRB type. Ruelle's response theory provides the tools to compute the change in the measure  $\mu(d\mathbf{x})$  of an autonomous Axiom A system caused by weak perturbations of intensity  $\epsilon$ , in terms of the unperturbed properties of the system. Thanks to the fact that the invariant measure of that system is differentiable with respect to  $\epsilon$ , even if supported on a strange attractor. Moving to the non-autonomous version of Ruelle's theory, it makes it possible to calculate the time-dependent measure  $\mu_t(d\mathbf{x})$  on the PBA by means of the computation of the time-dependent corrections to it with respect to a reference state. Providing a general formulation for the climate system's response to perturbations.

Assuming that the following may be written:

$$\dot{x} = F(x, t) = F(x) + \epsilon X(x, t), \quad (2.1)$$

where  $\forall t \in \mathbb{R}$  and  $\forall x \in \mathcal{Y} \subset \mathbb{R}^d$ ,  $|\epsilon X(x, t)| \ll |F(x)|$ . That way,  $F(x)$  can be taken as the background dynamics and  $\epsilon X(x, t)$  as the perturbation. They restrict the analysis to the separable case  $\dot{x} = F(x) + \epsilon X(x)T(t)$ , without losing generality.

Now, in order to evaluate the expectation value of a measurable observable  $\mathcal{O}(x)$ , i.e.  $\langle \mathcal{O} \rangle^\epsilon(t)$ , with respect to the measure  $\mu_t(dx)$  of the system, it is written:

$$\langle \mathcal{O} \rangle^\epsilon(t) = \int \mathcal{O}(x) \mu_t(dx) = \langle \mathcal{O} \rangle_0 + \sum_{j=1}^{\infty} \epsilon^j \langle \mathcal{O} \rangle_0^{(j)}(t) \quad (2.2)$$

where  $\langle \mathcal{O} \rangle_0 = \int \mathcal{O}(x) \bar{\mu}(dx)$  is the expectation value of the observable  $\mathcal{O}$  with respect to the SRB invariant measure  $\bar{\mu}(dx)$  of the autonomous dynamical system  $\dot{x} = F(x)$ . Then, they restrict themselves to the linear correction term, which can

be written as follows

$$\langle \mathcal{O} \rangle_0^{(1)}(t) = \int \int_0^\infty \Lambda \mathcal{S}_0^\tau \mathcal{O}(x) \mathbb{T}(t - \tau) d\tau \bar{\mu}(dx) = \int_0^\infty G_{\mathcal{O}, X}^{(1)}(\tau) \mathbb{T}(t - \tau) d\tau \quad (2.3)$$

where the Green's function  $G_{\mathcal{O}, X}^{(1)}(\tau)$  is given by

$$G_{\mathcal{O}, X}^{(1)}(\tau) = \int \Theta(\tau) \Lambda \mathcal{S}_0^\tau \mathcal{O}(x) \bar{\mu}(dx) \quad (2.4)$$

being  $\Lambda(\bullet) = X \cdot \nabla(\bullet)$ ,  $\mathcal{S}_0^\tau(\bullet) = \exp(tF \cdot \nabla)(\bullet)$  the semigroup of unperturbed Koopman operators.  $(\cdot)$  denotes the inner product in  $\mathcal{Y}$ , and the Heaviside distribution  $\Theta(\tau)$  enforces causality. If the unperturbed invariant measure  $\bar{\mu}(dx)$  is smooth with respect to the standard Lebesgue measure,  $\bar{\mu}(dx) = \tilde{\mu}(x)dx$  may be written, with  $\tilde{\mu}(x)$  the density and the Green's function can be rewritten as:

$$G_{\mathcal{O}, X}^{(1)}(\tau) = \Theta(\tau) \int \frac{-\nabla \cdot [\tilde{\mu}(x)X]}{\tilde{\mu}(x)} \mathcal{S}_0^\tau \mathcal{O}(x) \tilde{\mu}(x) dx = \Theta(\tau) \mathcal{C}(\Phi, \mathcal{S}_0^\tau \mathcal{O}) \quad (2.5)$$

where  $\Phi = \frac{-\nabla \cdot [\tilde{\mu}(x)X]}{\tilde{\mu}(x)}$  and  $\mathcal{C}(A, \mathcal{S}_0^\tau B)$  is the  $\tau$ -lagged correlation between the variables A and B. The average of  $\Phi$  vanishes. The Equation 2.5 is the appropriate generalization of the FDT for the non-autonomous and out-of-equilibrium system.

For a given time dependency of the forcing  $\mathbb{T}(t)$ , measuring the linear correction term  $\langle \mathcal{O} \rangle_0^{(1)}(t)$  from a set of experiments, the equations from Equation 2.1 to Equation 2.4 allow to derive the appropriate Green's function.

However, another approach to constructing the climate response to forcings is focusing on computing it directly from Equation 2.4. That way the problem of relying on the applicability of FDT is avoided (it fails in certain relevant cases). The difficulty in applying this direct approach comes from the fact that the formula contains contributions from both stable and unstable directions in the tangent space. And evaluating the contribution of the unstable directions is particularly hard. In several articles, like Lucarini et al. (2017), they proposed to evaluate the Green's function using an experimental but rigorous approach suggested by standard optics laboratory practice. The idea behind it is that using a set of selected probe experiments, normally, steplike increases of the parameter of interest, the Green's function could be constructed. And after that, using Equation 2.4 in combination to this obtained operator, the response of the system to a temporal pattern of interest for the forcing, could be predicted.

If a set of forced climate simulations and a background unperturbed one are given, this approach allows them to construct the Green's function's response operators for any desired observable. Hence, this tool kit allows one to treat a continuum of scenarios of temporal patterns forcings. The application that they present, consist in taking the set of equations that describe the unperturbed dynamics of climate

evolution in the form  $\dot{x} = F(x)$ , with the vector field  $X(x)$  as the 3D radiative forcing associated with the increment of  $CO_2$  concentration, and  $\epsilon T(t)$  its time modulation. Inserting  $T(t) = \epsilon \Theta(t)$  into Equation 2.4, for any climatic observable  $\mathcal{O}$ , it is obtained:

$$\frac{d}{dt} \langle \mathcal{O} \rangle_0^{(1)}(t) = \epsilon G_{\mathcal{O}, [CO_2]}^{(1)}(t) \quad (2.6)$$

They estimate  $\frac{d}{dt} \langle \mathcal{O} \rangle_0^{(1)}(t)$  taking the system's average of response over an ensemble of initial states and use the previous equation to derive their estimate of  $G_{\mathcal{O}, [CO_2]}^{(1)}(t)$  by means of assuming linearity in the response.

In the work presented in Lucarini et al. (2017), they show the results of applying the previous explained theory to an increase of  $CO_2$  concentration, finding good agreement between the ensemble average of 200 simulations using the model PlaSim and the prediction using the Green's function. Also, they show the results for a more localized prediction, with the outcome of predicting the change in the zonal averages of the surface temperature. Being it a good reproduction of the spatial patterns of temperature change as well.

## 2.3 Limitations of the Theory

Even if the application of Ruelle's linear response theory to climate change can provide an improvement upon the standard methods of forward integration of model ensembles with perturbed parameters and initial states, and applying also to systems out of thermodynamic equilibrium, it comes with some limitations.

The first limitation that they explain as well in Ghil and Lucarini (2019), given the inherent linearity of Ruelle's linear response theory, it is indeed limited to small perturbations in parameters. This limitation is overcome by the Exact Response Theory that will be studied later in the present work, as it is not restricted to linear cases.

As a second limitation of the theory, it was found in this work that the assumption of the climate system being an Axiom A system is not completely rigorous. As introduced before, Ruelle's linear response theory is developed for Axiom A systems. Being the definition of Axiom A dynamical systems:

Let  $M$  be a smooth manifold with a diffeomorphism  $f : M \rightarrow M$ . Then  $f$  is an axiom A diffeomorphism if the following two conditions hold:

1. The nonwandering set of  $f$ ,  $\Omega(f)$ , is a hyperbolic set and compact.
2. The set of periodic points of  $f$  is dense in  $\Omega(f)$ .

Thus, the hypothesis to consider the system Axiom A are too stringent. Since typically systems of physical interest are not ergodic, being Axiom A way out of

reach. It is true that analogy may be considered, if it proves some useful tool to use.

Lastly, as aforementioned, constructing directly the response operator using the Ruelle formula, Equation 2.3, is an arduous process given the difficulties associated to the contribution coming from both the unstable and stable directions. Therefore, they are forced to proceed with an experimental approach, deriving the Green function from a set of experiments, with the handicaps associated to that procedure, as avoiding the noise could be.

# Chapter 3

## Exact Response theory

### 3.1 Introduction

The Exact Response Theory that is going to be explained through this chapter was developed within the field of Nonequilibrium Molecular Dynamics. The theory introduces the Dissipation Function,  $\Omega$ , as its basis. A parallel could be drawn between this quantity and the thermodynamic potentials, as these potentials determine equilibrium state properties while the Dissipation Function determines nonequilibrium properties. The Dissipation Function can be used to determine the exact response of particle systems obeying classical mechanical laws, in the presence of perturbations of arbitrary size and modifications of states. It may also be used to express the response of a single system. As it will be explained later, the Dissipation Function can be associated with the entropy production rate. However, when studying general dynamical systems, like the case of the study conducted here, the function can still be defined leading to formal, thermodynamic-like relations. In that case, such relations may lack physical meaning, but they may establish interesting characterizations of the dynamics.

The Dissipation Function was first explicitly introduced in Evans and Searles (2002). After that, it was developed as the observable of interest in Fluctuation Relations in Evans et al. (2005) and Searles et al. (2007).

The Nonequilibrium Molecular Dynamics, NEMD, is a computational method born as a variation of Equilibrium Molecular Dynamics (when no dissipative forces are involved), MD. The nonequilibrium case was developed to model the nonequilibrium steady state that can be reached when the system is driven away from equilibrium by external forces that constantly feed energy in it, and, at the same time, it is in contact with some environment that removes energy from it. To do that, there were introduced the "synthetic forces", that is, forces that do not exist in nature, but that provide the way to transform a difficult boundary condition problem into a much simpler mechanical one. What they do in practice is constraining the dynamics of the particle systems in different ways.

They realized that the phase space contraction rate,  $-\Lambda$ , which is proportional to the thermostating term, could as well be proportional to the entropy production

when thermodynamics apply. In Searles and Evans (1999), the authors applied a representation of SRB measures to the fluctuations of  $\Lambda$  in a Gaussian isoenergetic shearing system, in that case the quantity  $\Lambda$  equals the entropy production, when local thermodynamic equilibrium holds. They derived and tested the Fluctuation Relation, which for a time reversal invariant model of shearing fluids, states that positive values of the energy dissipation, are exponentially more probable than the opposite. The later was interpreted as an explanation of the second law of thermodynamics and motivated a huge amount of activity. However, even if the phase space volumes variation rate,  $\Lambda$ , is easy to handle in dynamical systems theory, it was revealed that only in a limit set of cases it is justified the identification with the entropy production. In addition to that, in time dependent settings and when fluctuations are involved, the relation of  $\Lambda$  and the physically measurable properties results rather loose.

It was at that point when the Dissipation Function,  $\Omega$ , was presented as the physically relevant quantity for those cases where NEMD is applied. The Dissipation Function preserves the meaning of energy dissipation rate even when local thermodynamic equilibrium does not hold, and thermodynamic quantities such as the entropy production do not exist. Only in small amount of cases  $\Omega$  equals  $-\Lambda$ , even though steady states averages of both magnitudes may be simply related to each other. The Dissipation Function, as it will be shown later, can be used to solve in terms of physically relevant quantities the Liouville equation, and so, to investigate the exact average response of the observables of ensembles of systems. Moreover,  $\Omega$  may be used to provide conditions for the evolution of single systems. These results are expressed in terms of a physically measurable quantity, instead of being expressed in terms of an abstract phase space quantity.

## 3.2 Mathematical Framework

Considering the system of interest to be described by a dynamical system, whose equation of motion is

$$\dot{\mathbf{x}} = \mathbf{V}(\mathbf{x}); \quad x \in \mathcal{M} \subset \mathbb{R}^n$$

with  $\mathcal{M}$  the phase space. The solution at time  $t \in \mathbb{R}$  with initial condition  $\mathbf{x}$ , is given by  $\mathcal{S}^t \mathbf{x}$ . So that,  $\mathcal{S}^t$  is a flow on  $\mathcal{M}$ ,  $\mathcal{S}^t : \mathcal{M} \rightarrow \mathcal{M}$ .

Being  $\mu_0$  the probability measure absolutely continuous w.r.t. the Lebesgue measure  $d\mathbf{x}$ , such that  $d\mu_0(\mathbf{x}) = f_0(\mathbf{x})d\mathbf{x}$ ,  $f_0(\mathbf{x})$  corresponds to the density function, positive and continuously differentiable. In the event that  $\mu_0$  is not invariant, it evolves under the dynamics, so that at time  $t$  it is expressed by  $\mu_t(E) = \mu_0(\mathcal{S}^{-t}E)$ , for each measurable set  $E \subset \mathcal{M}$ .

The evolution of the density function  $f_0$  under the dynamics is described by the generalised Liouville equation:

$$\frac{\partial f_t}{\partial t}(\mathbf{x}) = -\nabla \cdot (f_t(\mathbf{x})\mathbf{V}(\mathbf{x})) = -\mathbf{V}(\mathbf{x}) \cdot \nabla f_t(\mathbf{x}) - f_t(\mathbf{x})\nabla \cdot \mathbf{V}(\mathbf{x})$$

Thus, the total derivative of  $f_t(\mathbf{x})$  can be expressed as

$$\frac{df_t}{dt}(\mathbf{x}) = \frac{\partial f_t}{\partial t}(\mathbf{x}) + \mathbf{V}(\mathbf{x}) \cdot \nabla f_t(\mathbf{x}) = -f_t(\mathbf{x}) \nabla \cdot \mathbf{V}(\mathbf{x})$$

in the Lagrangian form.

Introducing the phase-space expansion rate  $\Lambda$ , that denotes the phase space variation rate arising from the dynamics

$$\Lambda(\mathbf{x}) = \frac{d}{d\mathbf{x}} \cdot \dot{\mathbf{x}} = \nabla \cdot \mathbf{V}(\mathbf{x}),$$

whose negative is called the phase space contraction rate. Now, the partial and total derivatives of  $f_t$  with respect to time can be expressed in terms of this quantity

$$\frac{df_t}{dt}(\mathbf{x}) = -f_t(\mathbf{x})\Lambda(\mathbf{x}) \quad (3.1)$$

$$\frac{\partial f_t}{\partial t} = -f_t(\mathbf{x})\Lambda(\mathbf{x}) - \mathbf{V}(\mathbf{x}) \cdot \nabla f_t(\mathbf{x}) = -f_t(\mathbf{x})[\Lambda(\mathbf{x}) + \mathbf{V}(\mathbf{x}) \cdot \nabla \ln f_t(\mathbf{x})] \quad (3.2)$$

The so-called dissipation function corresponds to

$$\Omega^{f_t}(\mathbf{x}) := -[\Lambda(\mathbf{x}) + \mathbf{V}(\mathbf{x}) \cdot \nabla \ln f_t(\mathbf{x})] \quad (3.3)$$

The origin of the name dissipation function comes from the fact that this quantity corresponds, for a non-equilibrium molecular dynamics, to generalized entropy production, as aforementioned. In the field of dissipative dynamical systems, generally associated with the concept of diminishing volume of the occupied phase space over time, that corresponds to a negative phase-average of  $\Lambda$ . The notion of dissipation arises from considering the second law of thermodynamics applied to collections of macrostates. However, the dissipation function  $\Omega^f$  arises in a space of microstates, each of which affords the complete description of a macroscopic object at a given time.

The superscript  $f_t$  is used to explicitly denote the dependence on the density function  $f_t$ . That way it is obtained

$$\frac{\partial f_t}{\partial t} = -f_t(\mathbf{x})\Omega^{f_t}(\mathbf{x}), \quad (3.4)$$

the evolution of  $f_t$  sitting at the point  $\mathbf{x}$ , i.e. the Eulerian form. Here it is clear that the condition for the initial distribution to remain invariant under the flow is given by  $\Omega^{f_t}$ . If  $\Omega^{f_t} \equiv 0$ , the distribution at a point remains unchanged over time, which is a necessary and sufficient condition for the invariance of the probability density at  $\mathbf{x}$ , when  $f_0(\mathbf{x}) > 0$ .

While Equation 3.1 describes the evolution of  $f_t$  along the flow, exposing that the distribution does not change along the trajectory, if and only if  $\Lambda(\mathcal{S}^t \mathbf{x}) = 0$ .

Given any observable  $\mathcal{O} : \mathcal{M} \rightarrow \mathbb{R}$ , its ensemble average according to the measure  $d\mu_t(\mathbf{x}) = f_t(\mathbf{x})d\mathbf{x}$  is defined as

$$\langle \mathcal{O} \rangle_t = \int_{\mathcal{M}} \mathcal{O}(\mathbf{x}) f_t(\mathbf{x}) d\mathbf{x}. \quad (3.5)$$

In order to determine the evolution of the ensemble average of any observable, it is useful the evolution of  $f_t$  in terms of  $f_0$ . Thus, directly integrating Equation 3.1

$$f_{s+t}(\mathcal{S}^t \mathbf{x}) = \exp\{-\Lambda_{0,t}(\mathbf{x})\} f_s(\mathbf{x}) \quad (3.6)$$

where the following notation is used

$$\mathcal{O}_{r,s}(\mathbf{x}) = \int_r^s \mathcal{O}(\mathcal{S}^\tau \mathbf{x}) d\tau.$$

Nonetheless, Equation 3.6 does not provide  $f_t(\mathbf{x})$  in terms of  $f_0(\mathbf{x})$ , but provides  $f_t(\mathcal{S}^t \mathbf{x})$  in terms of  $f_0(\mathbf{x})$ . Thus, rather than that one, the equation used is Equation 3.4. To get to the final useful expression, the following relation is taken into account:

$$\begin{aligned} \Omega_{0,s}^{f_t}(\mathbf{x}) &= \int_0^s \Omega^{f_t}(\mathcal{S}^u \mathbf{x}) du = \int_0^s [\Lambda(\mathcal{S}^u \mathbf{x}) + \mathbf{V}(\mathcal{S}^u \mathbf{x}) \cdot \nabla \ln f_t(\mathcal{S}^u \mathbf{x})] du \\ &= -\Lambda_{0,s}(\mathbf{x}) - \int_0^s \frac{d}{du} \ln f_t(\mathcal{S}^u \mathbf{x}) du \\ &= -\Lambda_{0,s}(\mathbf{x}) - \frac{\ln f_t(\mathcal{S}^s \mathbf{x})}{f_t(\mathbf{x})} \\ &= \ln \frac{f_t(\mathbf{x})}{f_t(\mathcal{S}^s \mathbf{x})} - \Lambda_{0,s}(\mathbf{x}) \end{aligned}$$

where it has been used the equality  $\mathbf{V}(\mathcal{S}^u \mathbf{x}) \cdot \nabla A(\mathcal{S}^u \mathbf{x}) = \frac{d}{du} A(\mathcal{S}^u \mathbf{x})$ , valid when  $A$  has no explicit dependence on  $u$ , as in the case of  $\ln f_t(\mathcal{S}^u \mathbf{x})$ .

Now, adding the exponential and  $f_s(\mathcal{S}^{s+t} \mathbf{x})$  to the last expression

$$\begin{aligned} \exp\{\Omega_{s,s+t}^{f_s}(\mathbf{x})\} f_s(\mathcal{S}^{s+t} \mathbf{x}) &= \left( \frac{f_s(\mathcal{S}^s \mathbf{x})}{f_s(\mathcal{S}^{s+t} \mathbf{x})} \exp\{-\Lambda_{s,s+t}(\mathbf{x})\} \right) f_s(\mathcal{S}^{s+t} \mathbf{x}) \\ &= \exp\{-\Lambda_{s,s+t}(\mathbf{x})\} f_s(\mathcal{S}^s \mathbf{x}) = f_{s+t}(\mathcal{S}^{s+t} \mathbf{x}) \end{aligned}$$

It leads to the searched expression, relating  $f_t(\mathbf{x})$  in terms of  $f_0(\mathbf{x})$ ,

$$f_{s+t}(\mathbf{x}) = \exp\{\Omega_{-t,0}^{f_s}(\mathbf{x})\} f_s(\mathbf{x}) \quad (3.7)$$

Where it is implicitly assumed that the dynamics is invertible (existence of the unique projection of the dynamics backwards in time), given that the distribution function at future times,  $\Omega^{f_s}$ , is integrated forward in time over trajectories that begin in the past.

Therefore, an expression for  $\langle \mathcal{O} \rangle_{f_t}$  could be written in terms of  $\Omega^{f_0}$  and  $f_t$ . In order to obtain the expression, the following two identities are used

$$\begin{aligned} \mathcal{O}_{0,s}(\mathbf{x}) &= \int_0^s \mathcal{O}(\mathcal{S}^u \mathbf{x}) du = \int_\tau^{s+\tau} \mathcal{O}(\mathcal{S}^{u-\tau} \mathbf{x}) du \\ &= \int_\tau^{s+\tau} \mathcal{O}(\mathcal{S}^{-\tau} \mathcal{S}^u \mathbf{x}) du = \mathcal{O}_{\tau,s+\tau}(\mathcal{S}^{-\tau} \mathbf{x}) ; \end{aligned}$$

$$\begin{aligned} \langle \mathcal{O} \rangle_{t+s} &= \int \mathcal{O}(\mathbf{x}) f_{t+s}(\mathbf{x}) d\mathbf{x} = \int \mathcal{O}(\mathcal{S}^s(\mathcal{S}^{-s} \mathbf{x})) f_{t+s}(\mathcal{S}^s(\mathcal{S}^{-s} \mathbf{x})) \left| \frac{\partial \mathbf{x}}{\partial \mathcal{S}^{-s} \mathbf{x}} \right| d(\mathcal{S}^{-s} \mathbf{x}) \\ &= \int \mathcal{O}(\mathcal{S}^s(\mathcal{S}^{-s} \mathbf{x})) f_{t+s}(\mathcal{S}^s(\mathcal{S}^{-s} \mathbf{x})) \exp\{\Lambda_{-s,0}(\mathbf{x})\} d(\mathcal{S}^{-s} \mathbf{x}) \\ &= \int \mathcal{O}(\mathcal{S}^s(\mathcal{S}^{-s} \mathbf{x})) f_{t+s}(\mathcal{S}^s(\mathcal{S}^{-s} \mathbf{x})) \exp\{\Lambda_{0,s}(\mathcal{S}^{-s} \mathbf{x})\} d(\mathcal{S}^{-s} \mathbf{x}) \\ &= \int \mathcal{O}(\mathcal{S}^s \mathbf{x}) f_{t+s}(\mathcal{S}^s \mathbf{x}) \exp\{\Lambda_{0,s}(\mathbf{x})\} d(\mathbf{x}) \\ &= \int \mathcal{O}(\mathcal{S}^s \mathbf{x}) f_t(\mathbf{x}) d(\mathbf{x}) = \langle \mathcal{O} \circ \mathcal{S}^s \rangle_t. \end{aligned}$$

That way, the last expression provides:

$$\langle \mathcal{O} \rangle_{t+s} = \int \mathcal{O}(\mathcal{S}^s \mathbf{x}) f_t(\mathbf{x}) d(\mathbf{x}) = \langle \mathcal{O} \circ \mathcal{S}^s \rangle_t. \quad (3.8)$$

Moreover, for any arbitrary observable, it is observed,

$$\begin{aligned} \frac{d}{ds} \langle \mathcal{O} \rangle_{f_s} &= \lim_{h \rightarrow 0} \frac{1}{h} [\langle \mathcal{O} \rangle_{f_{s+h}} - \langle \mathcal{O} \rangle_{f_s}] \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \int [\mathcal{O}(\mathbf{x}) f_{s+h}(\mathbf{x}) - \mathcal{O}(\mathbf{x}) f_s(\mathbf{x})] d\mathbf{x} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \int \mathcal{O}(\mathbf{x}) [f_r(\mathbf{x}) \exp\{\Omega_{r-s-h,0}^{f_r}(\mathbf{x})\} - f_r(\mathbf{x}) \exp\{\Omega_{r-s,0}^{f_r}(\mathbf{x})\}] d\mathbf{x} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \int \mathcal{O}(\mathbf{x}) f_r(\mathbf{x}) \exp\{\Omega_{r-s,0}^{f_r}(\mathbf{x})\} [\exp\{\Omega_{r-s-h,r-s}^{f_r}(\mathbf{x})\} - 1] d\mathbf{x} \end{aligned}$$

where the Equation 3.7 is used.

Now, studying the part of the integral that depends on  $h$ ,

$$\begin{aligned}
\lim_{h \rightarrow 0} \frac{1}{h} [\exp\{\Omega_{r-s-h, r-s}^{f_r}(\mathbf{x})\} - 1] &= \lim_{h \rightarrow 0} \frac{1}{h} [\exp\{\int_{r-s-h}^{r-s} \Omega^{f_r}(\mathcal{S}^u \mathbf{x}) du\} - 1] \\
&= \frac{d}{dh} \exp\{\int_{r-s-h}^{r-s} \Omega^{f_r}(\mathcal{S}^u \mathbf{x}) du\} |_{h=0} \\
&= \Omega^{f_r}(\mathcal{S}^{r-s-h} \mathbf{x}) \exp\{\int_{r-s-h}^{r-s} \Omega^{f_r}(\mathcal{S}^h \mathbf{x})\} |_{h=0} \\
&= \Omega^{f_r}(\mathcal{S}^{r-s} \mathbf{x})
\end{aligned}$$

Hence, one may write

$$\begin{aligned}
\frac{d}{ds} \langle \mathcal{O} \rangle_{f_s} &= \int \mathcal{O}(\mathbf{x}) f_r(\mathbf{x}) \exp\{\Omega_{r-s,0}^{f_r}(\mathbf{x})\} \Omega^{f_r}(\mathcal{S}^{r-s} \mathbf{x}) d\mathbf{x} \\
&= \int \mathcal{O}(\mathbf{x}) f_s(\mathbf{x}) \Omega^{f_r}(\mathcal{S}^{r-s} \mathbf{x}) d\mathbf{x} = \langle \mathcal{O} \cdot (\Omega^{f_r} \circ \mathcal{S}^{r-s}) \rangle_{f_s}.
\end{aligned}$$

using again Equation 3.7. One last useful equality is:

$$\frac{d}{ds} \langle \mathcal{O} \rangle_{f_s} = \langle \mathcal{O} \cdot (\Omega^{f_r} \circ \mathcal{S}^{r-s}) \rangle_{f_s}, \quad (3.9)$$

which in the case  $r = 0$  would be, using Equation 3.8,

$$\frac{d}{ds} \langle \mathcal{O} \rangle_{f_s} = \langle \mathcal{O} \cdot (\Omega^{f_0} \circ \mathcal{S}^{-s}) \rangle_{f_s} = \langle (\mathcal{O} \circ \mathcal{S}^s) \cdot \Omega^{f_0} \rangle_{f_0} \quad (3.10)$$

While if  $r = s$ ,

$$\frac{d}{ds} \langle \mathcal{O} \rangle_{f_s} = \langle \mathcal{O} \cdot \Omega^{f_s} \rangle_{f_s} \quad (3.11)$$

Thus, the evolution of  $\langle \mathcal{O} \rangle_{f_s}$  in terms of averages with respect to the initial distribution  $f_0$  is described by

$$\langle \mathcal{O} \rangle_{f_t} = \langle \mathcal{O} \rangle_{f_0} + \int_0^t \langle (\mathcal{O} \circ \mathcal{S}^s) \cdot \Omega^{f_0} \rangle_{f_0} ds \quad (3.12)$$

The Equation 3.12 is the one that allows one to compute the response of the system to a perturbation, and so, it will be computed later in the work.

### 3.2.1 Conditions on the Dynamics and Probability Density Function

As explained above, the dissipation function can be obtained from Equation 3.2, resulting in Equation 3.4. This operation can be performed under the following standard conditions:

1. The vector field  $\mathbf{V}(\mathbf{x})$  should be everywhere differentiable in the phase space  $\mathcal{M}$  for  $\Lambda$  to exist.

2. The initial probability density function  $f_0$  should be everywhere positive in  $\mathcal{M}$  for its logarithm to exist.
3. The initial probability density function  $f_0$  should be everywhere differentiable in  $\mathcal{M}$  for its gradient to exist.

These conditions also guarantee the continuity of the dissipation function  $\Omega$ , which is used to obtain Equation 3.12 from Equation 3.5, via Equation 3.7.

### 3.3 Current Applications of Exact Response Theory

Previous works where the exact response theory is applied can be found. In particular, the work developed in Amadori et al. (2022) was considered useful to expose in a clear way the application of the theory. Thus, a summary of it is provided this section.

#### 3.3.1 Dynamics of Kuramoto Oscillators

In Amadori et al. (2022) the response theory is applied to the dynamics of the Kuramoto oscillators. The Kuramoto model is contemplated as a prototype of many particle systems exhibiting synchronization. It is a dissipative system with many degrees of freedom undergoing nonequilibrium phase transitions.

As Kuramoto dynamics undergo synchronization transitions, the linear response theory could not apply. Indeed, it is showed through a comparison between the application of the exact response theory and the linear response theory, that the later one fails while the exact one is capable of handling small perturbations which may result in large modification of the state.

In the work it is obtained the synchronization within the formalism of the Dissipation Function, thus showing how such a behavior is captured by the exact response theory, being not evidenced by the linear theory. Synchronization corresponds indeed to the maximum value of the Dissipation Function, which they prove to be attained in time.

#### Properties of Kuramoto System

As introduced in Amadori et al. (2022), the following set of coupled first order ODEs, define the Kuramoto dynamics:

$$\dot{\theta}_i(t) = \omega_i + \frac{K}{N} \sum_{j=1}^N \sin(\theta_j(t) - \theta_i(t)), \quad i = 1, \dots, N. \quad (3.13)$$

It is defined on the N-dimensional torus,  $\mathcal{T}^N = (\mathbb{R}/(2\pi\mathbb{Z}))^N$ , where  $N \geq 1$  is the number of oscillators,  $K > 0$  is a constant, and the natural frequencies  $\omega_i \in \mathbb{R}$  are drawn for some given distribution  $g(\omega)$ .

The oscillators can be represented by points rotating on the unit circle centered at

the origin of the complex plane, so mathematically  $e^{i\theta_j}$  with  $j = 1, \dots, N$ . Now, introducing the polar coordinates of the barycenter,

$$Re^{i\Phi} = \frac{1}{N} \sum_{j=1}^N e^{i\theta_j}, \quad (3.14)$$

with  $R \in [0, 1]$  and  $\Phi \in \mathbb{R}$  (defined if  $R > 0$ ), the Equation 3.13 can be rewritten to:

$$\dot{\theta}_i(t) = \omega_i + KR \sin(\Phi - \theta_i), \quad i = 1, \dots, N. \quad (3.15)$$

$R = R(\theta(t))$  is the order parameter and  $\Phi = \Phi(\theta(t))$  is the collective phase, with the phase  $\theta = (\theta_1, \dots, \theta_N) \in \mathcal{M} = \mathcal{T}^N$ , where  $\mathcal{M}$  denotes the phase space.

As they point out in Amadori et al. (2022), "*a complete frequency synchronization occurs when the differences  $\theta_i(t) - \theta_j(t)$  tend to a constant for all  $i$  and  $j$ , and  $R(\theta(t))$  tends to a given  $R^\infty \in (0, 1]$ , as  $t \rightarrow +\infty$ . In case  $R^\infty = 1$ , all the  $N$  terms of the sum in Equation 3.14 coincide, hence the Kuramoto system undergoes phase synchronization*".

From identities that Equation 3.14 provides, they find the following relation,

$$R^2 = \frac{1}{N^2} \sum_{i,j=1}^N \cos(\theta_j - \theta_i). \quad (3.16)$$

Further rewriting of Equation 3.15 leads to

$$\dot{\theta} = W + V(\theta) = V_K(\theta) \quad (3.17)$$

where  $W = (\omega_1, \dots, \omega_N)$  is interpreted as "*an equilibrium vector field of  $N$  natural frequencies*", whereas  $V$  represents "*a non equilibrium perturbation*", composed of:

$$V_i(\theta) = \frac{K}{N} \sum_{j=1}^N \sin(\theta_j - \theta_i) = KR \sin(\Phi - \theta_i), \quad i = 1, \dots, N. \quad (3.18)$$

The phase space volumes variation rate is then given by the equation:

$$\Lambda = K(1 - NR^2) \quad (3.19)$$

### Applying Response Theory

Focusing on identical oscillators, that is, where for all the natural frequencies  $\omega_i = \omega$ . The following dynamics are proposed:

$$\dot{\theta} = \begin{cases} W & t < 0 \\ W + V(\theta) & t > 0 \end{cases} \quad (3.20)$$

where  $V_0(\theta) = W = (\omega, \dots, \omega)$  represents the unperturbed dynamics, which corresponds to  $K = 0$ . This dynamics are conservative, given that  $\text{div}_\theta V_0 = 0$ , thus, the

corresponding steady state could be considered an equilibrium state. However, at time  $t = 0$ , the perturbation  $V(\theta)$  is switched on, leading to the Kuramoto dynamics (non-conservative) described above.

As initial probability density, they take  $f_0(\theta) = 2\pi^{-N}$ , which is invariant under the unperturbed dynamics. Leading to  $\Omega_V^{f_0} \equiv 0$  (for the unperturbed dynamics), while using the perturbed system the Dissipation function has the form:

$$\Omega_V^{f_0} = K(NR^2 - 1) = \frac{K}{N} \sum_{i,j=1}^N \cos(\theta_j - \theta_i) - K \quad (3.21)$$

Now, if the Equation 3.12 is applied taking as observable the Dissipation Function, so  $\mathcal{O} = \Omega_V^{f_0}$ , the response is computed as follows:

$$\langle \Omega_V^{f_0} \rangle_{f_t} = \langle \Omega_V^{f_0} \rangle_{f_0} + \int_0^t \langle (\Omega_V^{f_0} \circ \mathcal{S}^s) \cdot \Omega_V^{f_0} \rangle_{f_0} ds. \quad (3.22)$$

Knowing that,  $\langle R^2 \rangle_0 = 1/N$ , the phase space average  $\langle \Omega_V^{f_0} \rangle_{f_0}$  vanishes. Hence, only the second integral must be computed. Substituting Equation 3.21 into Equation 3.22, it is obtained:

$$\langle \Omega_V^{f_0} \rangle_{f_t} = K^2 N^2 \int_0^t \langle R^2 [R^2 \circ \mathcal{S}^s] \rangle_{f_0} ds - K^2 N \int_0^t \langle R^2 \circ \mathcal{S}^s \rangle_{f_0} ds \quad (3.23)$$

If the general case is reduced to two oscillators, i.e.  $N = 2$ , explicit calculations can be carried out. In particular they get to the following relations:

$$\begin{aligned} \langle R^2 \circ \mathcal{S}^s \rangle_{f_0} &= \frac{1}{(2\pi)^2} \int_{\mathcal{M}} \frac{1}{\tan^2(\frac{\theta_1 - \theta_2}{2}) e^{-2Ks} + 1} d\theta = \frac{1}{e^{-Ks} + 1}, \\ \langle R^2 (R^2 \circ \mathcal{S}^s) \rangle_{f_0} &= \frac{1}{8\pi^2} \int_{\mathcal{M}} \frac{1 + \cos(\theta_1 - \theta_2)}{\tan^2(\frac{\theta_1 - \theta_2}{2}) e^{-2Ks} + 1} d\theta = \frac{2e^{-Ks} + 1}{2(e^{-Ks} + 1)^2}. \end{aligned}$$

Therefore, integrating both of them through time the following expressions are found

$$\int_0^t \langle R^2 \circ \mathcal{S}^s \rangle_{f_0} ds = t + \frac{\ln(e^{-Kt} + 1)}{K} - \frac{\ln 2}{K},$$

$$\int_0^t \langle R^2 (R^2 \circ \mathcal{S}^s) \rangle_{f_0} ds = \frac{t}{2} + \frac{1}{2K} \left[ \frac{3}{2} + \ln \left( \frac{e^{-Kt} + 1}{2} \right) - \frac{2}{e^{Kt} + 1} - \frac{1}{e^{-Kt} + 1} \right].$$

Then, finally, the following explicit expressions are obtained

$$\langle \Omega_V^{f_0} \rangle_{f_t} = K \tanh \left( \frac{Kt}{2} \right), \quad (3.24)$$

$$\langle (\Omega_V^{f_0} \circ \mathcal{S}^s) \cdot \Omega_V^{f_0} \rangle_{f_0} = \frac{K^2}{1 + \cosh(Kt)}. \quad (3.25)$$

Apart from this analytical results, in the article they compute the evolution of  $\Omega$  numerically.

This numerical computation is replicated here, as part of the learning phase of the theory, following the procedure explained in chapter 6. The results are showed in Figure 3.1.

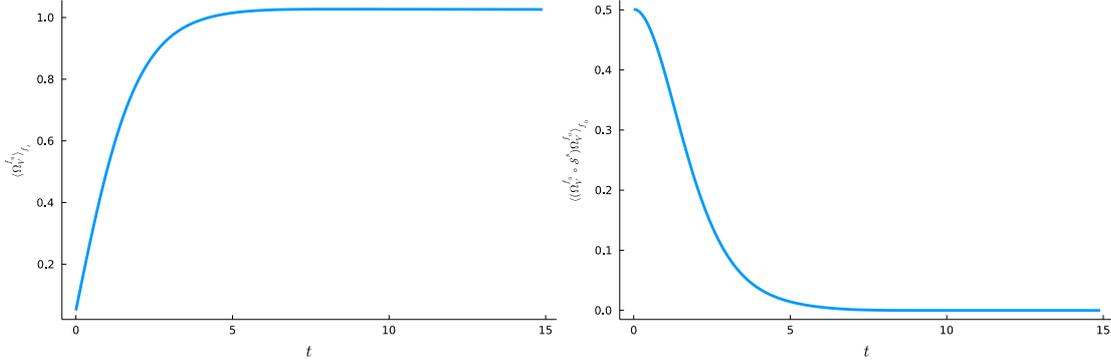


Figure 3.1: Evolution of  $\langle \Omega_V^{f_0} \rangle_{f_t}$  and  $\langle (\Omega_V^{f_0} \circ \mathcal{S}^s) \Omega_V^{f_0} \rangle_{f_0}$  with the time when  $N = 2$ ,  $K = 1$  and  $\omega = 0$ . The averages were taken over a set of 5041 trajectories with initial data sampled from the uniform distribution  $[0, 2\pi)$ .

It was proven that the average of  $\Omega$  tends to the value  $K$  when  $t \rightarrow \infty$ . As well as the showed case when  $N = 2$ , in Amadori et al. (2022) the numerical computation of the evolution of averaged  $\Omega$  is performed for the general case. When the number of oscillators  $N$  is large, the maximum value was found to be proportional to  $K$ , in particular  $\Omega$  tends to  $(N - 1)K$ .

### Comparison with Linear Response Theory

Later in the work, they compare the results given by the exact response theory with those of the linear response theory using a generalized Green-Kubo formula, as follows. Considering a perturbed vector field, defined as:

$$V_\varepsilon(\theta) = V_0(\theta) + \varepsilon V_p(\theta), \quad (3.26)$$

where the parameter  $\varepsilon$  represents the strength of the perturbations. Applying it to the before developed application to identical oscillators,  $\varepsilon$  is identified as  $K$ , and it is defined:

$$V_0(\theta) = \omega, \\ V_{p,j}(\theta) = R \sin(\Phi - \theta_j), \quad j = 1, \dots, N.$$

They denote  $\mathcal{S}_\varepsilon^t$  and  $\mathcal{S}_0^t$  the perturbed and unperturbed flows, respectively. It is obtained,

$$\Omega_\varepsilon^{f_0} = \Omega_0^{f_0} + \varepsilon \Omega_p^{f_0} = \varepsilon \Omega_p^{f_0},$$

where  $\Omega_0^{f_0}$  and  $\Omega_p^{f_0}$  correspond to the Dissipation Function evaluated in terms of the vector fields  $V_0$  and  $V_p$ , respectively. Then, the exact response is written as:

$$\langle \mathcal{O} \rangle_{t,\varepsilon} = \langle \mathcal{O} \rangle_0 + \varepsilon \int_0^t \langle (\mathcal{O} \circ \mathcal{S}_\varepsilon^\tau) \Omega_p^{f_0} \rangle_0 d\tau. \quad (3.27)$$

Since the formula is exact,  $\varepsilon$  does not need to be small, also it appears both as a factor multiplying the integral and in the perturbed flow  $\mathcal{S}_\varepsilon^t$ .

Whereas the linear response result is written as follows:

$$\langle \bar{\mathcal{O}} \rangle_{t,\varepsilon} = \int_{\mathcal{M}} \mathcal{O}(\theta) \bar{f}_t(\theta; \varepsilon) d\theta = \langle \mathcal{O} \rangle_0 + \varepsilon \int_0^t \langle (\mathcal{O} \circ \mathcal{S}_0^\tau) \Omega_p^{f_0} \rangle_0 d\tau. \quad (3.28)$$

Comparing them it is obtained:

$$\langle \mathcal{O} \rangle_{t,\varepsilon} - \langle \bar{\mathcal{O}} \rangle_{t,\varepsilon} = \varepsilon \int_0^t \langle [(\mathcal{O} \circ \mathcal{S}_\varepsilon^\tau) - (\mathcal{O} \circ \mathcal{S}_0^\tau)] \Omega_p^{f_0} \rangle_0 d\tau, \quad (3.29)$$

which shows that indeed, the two formulae tend to be the same, when small  $\varepsilon$  limit.

Now taking the observable as  $\mathcal{O} = \Omega_\varepsilon^{f_0} = \varepsilon \Omega_p^{f_0}$ , they find for  $N = 2$ :

$$\langle \Omega_\varepsilon^{f_0} \rangle_{t,\varepsilon} = \varepsilon \tanh\left(\frac{\varepsilon t}{2}\right), \quad \langle \bar{\Omega}_\varepsilon^{f_0} \rangle_{t,\varepsilon} = \frac{\varepsilon^2 t}{2},$$

which makes the difference:

$$\langle \Omega_\varepsilon^{f_0} \rangle_{t,\varepsilon} = \langle \bar{\Omega}_\varepsilon^{f_0} \rangle_{t,\varepsilon} + o(\varepsilon^2)t,$$

showing that for small times the difference between both responses is small, but diverges linearly as time grows.



# Chapter 4

## Dynamics of the System

### 4.1 Definition of the System

The chosen dynamical system to apply the exact response theory, is the model proposed in Saltzman and Maasch (1988), which describes the Pleistocene ice ages. The model considers three "slow response prognostic" variables, believed to be of prime relevance:

- Global Ice Mass,  $I$
- Atmospheric Carbon Dioxide,  $\mu$
- A measure of the strength of the oceanic " $CO_2$  pump" (believed to be related to the magnitude of North Atlantic Deep Water,  $N$ )

There are two additional diagnostic variables, which are determined by the mentioned "prognostic" variables.

- Mean Oceanic Surface Temperature,  $\tau$
- Permanent Sea Ice Extent,  $\eta$

At that time, there was an understanding that the control of the natural atmospheric  $CO_2$  variations resided in the state of the upper layer of the world ocean, as influenced by the circulation and mass exchanges with the deeper levels. When considering it at a basic level, thus, neglecting inputs due to volcanism, anthropogenic sources and the exchange with the continental biosphere (apparently minor), the atmosphere  $CO_2$  increases if the flux of  $CO_2$  from oceanic sources (in particular from warm low-latitude waters) exceeds the oceanic uptake (mainly in cold high-latitude waters). The balance described, is, said by Saltzman in Saltzman and Maasch (1988), "*highly vulnerable to changes in ocean circulation and mixing and associated changes in the chemical-biological-thermal state of the upper layer waters with which the atmosphere tends to (but probably never does) equilibrate*".

Based on some studies of that period, they assume that the primary controls of this delicate balance are the following, corresponding to or related to the above variables:

- **Sea surface temperature.** Colder water presents, even if small when considered alone, an effect of higher solubility and uptake of  $CO_2$ .
- **Permanent ice coverage as measured by summer sea ice extent.**
  - Large ice extent enhances the efficiency of high-latitude bioproductivity, by moving phytoplankton to lower latitudes. This effect is considered to be small.
  - Enhanced surface seasonal meltwater volume coming from sea ice fields provides a more stable cap on high-latitude waters. The most relevant effects is that it tends to inhibit deep convective mixing in high latitudes that can cause high  $CO_2$  levels in polar waters.
  - Large ice extent towards the equator increases the atmospheric and oceanic baroclinicity leading to: large horizontal surface water exchange between low and high latitudes, decreasing atmospheric  $CO_2$ ; increase precipitation over high-latitude water amplifying the effect of the meltwater described above, as it reduces the salinity; increase in shallow mechanical mixing in high-latitude surfaces which results in higher flux of  $CO_2$  from the atmosphere to the ocean.
- **The strength of the main thermohaline circulation, in particular the production of North Atlantic Deep Water (NADW).**
  - First order effects of a large mass of NADW in the ocean would be a decrease of  $CO_2$  in the atmosphere due to: stabilization of the ocean by filling the deeper levels with dense saline water, which prevents the convective overturnings that would bring up water carbon-rich water and increase the  $CO_2$  in the atmosphere; strong downwelling of  $CO_2$  in the NADW production zone and horizontal replacement with carbon-poor waters coming from lower latitudes, which translates in a higher solubility. Also, the transportation of nutrients throughout the world ocean by the NADW, might increase global productivity and  $CO_2$  downdraw.
  - However, a strong intrusion of relatively warmer water under cold Antarctic surface waters, destabilise deep convection, lowering the first order effects above mentioned. Hence, small increases of NADW from an equilibrium state tend to lower  $CO_2$ , but bigger ones could start increasing surface-deep vertical mixing, increasing  $CO_2$ . The authors in the paper adopted this hypothesis as one that could account for the asymmetry of the climatic response including the rapid changes of  $CO_2$ , NADW and ice mass during deglaciation. Nonetheless, from another, more generic, point of view, knowing that the fluxes due to deep convection overturnings are

a “highly nonlinear function of the density stratification, bifurcating to a relatively high value when a critical point of instability is approached or exceeded”, it could be expected that a negative anomaly of NADW produces a larger flux increase than the decrease in flux that a positive anomaly of the same magnitude generates.

- **Sea level change associated with global ice mass changes.**
  - Rising sea level comes with coral reef growth releasing excess  $CO_2$  to the surface waters and hence, to the atmosphere according to the corresponding reaction which reverse is operative during falling sea level.
  - During falling sea levels, nutrients stored in organic detritus which were deposited in the continental shelves during the preceding sea level rise, are now eroded back into the ocean, enhancing productivity and carbon burial in deeper ocean.
  - A small effect arises due to generation of vegetation sinks of  $CO_2$  when land masses in tropical areas are exposed after the fall in sea level.
- **Negative feedbacks.** Like in all natural systems, strong departures from equilibrium are opposed by dissipative effects. Some examples could be the increase in continental vegetation or augmented dissolution in surface water when higher levels of atmospheric  $CO_2$  are reached. Hence, they speculate that when the entire climatic system is far from equilibrium, this damping of  $CO_2$  extremes is maximized.

Therefore, after examining the implications of the above phenomena, the rate of change of atmospheric  $CO_2$  concentration,  $\mu$ , can be formulated in the following way:

$$\frac{d\mu}{dt} = \dot{\mu} = r_1\tau' - r_2\eta' - (r_3 - b_3N')N' - r_5\dot{I}' - (r_4 + b_4N'^2)\mu' + \mathcal{F}_\mu, \quad (4.1)$$

where,  $( )'$  means departure of the variable from equilibrium.  $\tau$  is global temperature of the water surface.  $\eta$  is global mean extent of permanent sea ice.  $I$  is global ice mass.  $N$  is the amount and extent of NADW. The parameters  $r_1, r_2, r_3, r_4, r_5, b_3$  and  $b_4$  are assumed to be positive rate constants and  $\mathcal{F}_\mu$  denotes external forcing due to direct inputs of  $CO_2$  (for example, volcanic effects unbalanced by weathering).

Now, the expressions of the rest of the variables are:

$$\dot{I}' = -s_1\tau' - s_2\mu' + s_3\eta' - s_4I' + \mathcal{F}_I, \quad (4.2)$$

$$\dot{N}' = -c_0I' - c_2N' + \mathcal{F}_N, \quad (4.3)$$

where  $s_1, s_2, s_3, s_4, c_0,$  and  $c_2$  are assumed to be positive rate constants and  $\mathcal{F}_I, \mathcal{F}_N$  denote external forcing.

It is further assumed that the fast response variables  $\tau$  and  $\eta$  can be expressed in terms of prescribed slow response variables  $I$ ,  $\mu$  and  $N$ :

$$\tau' = -\alpha I' + \beta \mu' + \mathcal{F}_\tau, \quad (4.4)$$

$$\eta' = -e_I I' + e_\mu \mu' + \mathcal{F}_\eta, \quad (4.5)$$

where  $\alpha$ ,  $\beta$ ,  $e_I$  and  $e_\mu$  are the equilibrium sensitivity coefficients assumed to be positive. While  $\mathcal{F}_\tau$  and  $\mathcal{F}_\eta$  represent the effects of external forcing, like Earth orbital could be. The possible dependence of  $\tau$  and  $\eta$  on  $N$  is uncertain, but neglected here.

Therefore, substituting Equation 4.4 and Equation 4.5 in Equation 4.1, Equation 4.2 and Equation 4.3 they obtain the following system governing the slow response variables:

$$\dot{I}' = -a_0 I' - a_1 \mu' + F_I \quad (4.6)$$

$$\dot{\mu}' = -b_0 I' + b_1 \mu' - (r_3 - b_3 N') N' - b_4 N'^2 \mu' + F_\mu \quad (4.7)$$

$$\dot{N}' = -c_0 I' - c_2 N' + F_N \quad (4.8)$$

where  $a_0 = (s_4 - s_1 \alpha - s_3 e_I)$ ,  $a_1 = (s_1 \beta + s_3 e_\mu + s_2)$ ,  $b_0 = (r_1 \alpha + r_2 e_I - r_5 a_0)$ ,  $b_1 = (r_1 \beta + r_2 e_\mu + r_5 a_1 - r_4)$ ,  $F_I = (\mathcal{F}_I - s_1 \mathcal{F}_\tau + s_3 \mathcal{F}_\eta)$ ,  $F_\mu = (\mathcal{F}_\mu + r_1 \mathcal{F}_\tau - r_2 \mathcal{F}_\eta - r_5 \mathcal{F}_I)$ , and  $F_N = \mathcal{F}_N$ .

The appropriate value of the rate constants in these equations is not known, however, after some assumptions and simplifications combined with the search of obtaining the behavior of the Vostok ice core measurements seen in Barnola et al. (1987) (giving atmospheric  $CO_2$ ) and recent  $\delta^{13}C$  measurements in Curry and Crowley (1987) (providing global ice mass), the values of this parameters are obtained.

In particular, if the above equations are rescaled with the transformations:  $t = [a_0^{-1}]t^*$ ,  $I' = [c_2 c_0^{-1} (a_0/b_4)^{1/2}]X$ ,  $\mu' = [c_2 (a_1 c_0)^{-1} (a_0^3/b_4)^{1/2}]Y$ , and  $N' = [(a_0/b_4)^{1/2}]Z$ , it is obtained

$$\begin{aligned} \dot{X} &= -X - Y \\ \dot{Y} &= -pZ + rY + sZ^2 - Z^2Y \\ \dot{Z} &= -q(X + Z) \end{aligned}$$

where  $(\dot{\phantom{x}}) = \frac{d(\phantom{x})}{dt^*}$ ,  $p = a_1 c_0 b_2 / a_0^2 c_2$ ,  $q = c_2 / a_0$ ,  $r = b_1 / a_0$  and  $s = a_1 b_3 c_0 (b_4 / a_0)^{1/2} / a_0 b_4 c_2$ . And, in particular, they took,  $p = 0.9$ ,  $q = 1.2$ ,  $r = 0.8$  and  $s = 0.8$ , obtaining the oscillatory response of  $\sim 115$  kyr, showed in Figure 4.1. In Saltzman and Maasch (1988) the figure is described as “*Time-dependent periodic solution for departures from equilibrium, in nondimensional units scaled to a range of unity, for an arbitrary 500-kyr period. Upper panel: global ice mass (solid) compared with SPECMAP*”

$\delta^{18}O$  curve (dashed) over past 200 kyr. Middle panel: atmospheric carbon dioxide (solid), compared with Vostok  $CO_2$  curve (dashed) over past 160 kyr. Lower panel: North Atlantic deep water. The vertical line is assumed to correspond to conditions at the present interglacial time”.

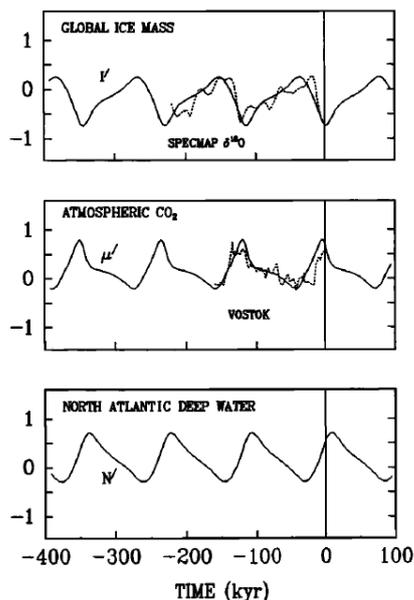


Figure 4.1: Solution of the system as showed in Saltzman and Maasch (1988).

## 4.2 Dynamics of the Resulting System

The expression for  $\mathbf{V}_0(\mathbf{x})$  that is going to be used in this study is the scalated one provided in Saltzman and Maasch (1988) and presented in section 4.1:

$$\dot{X} = -X - Y \quad (4.9)$$

$$\dot{Y} = -pZ + rY + sZ^2 - Z^2Y \quad (4.10)$$

$$\dot{Z} = -q(X + Z) \quad (4.11)$$

As a first approach, the evolution of an initial point can be simulated. In the Figure 4.2, such evolution through time is depicted for the three variables. Obtaining a similar figure as Figure 4.1 from Saltzman and Maasch (1988).

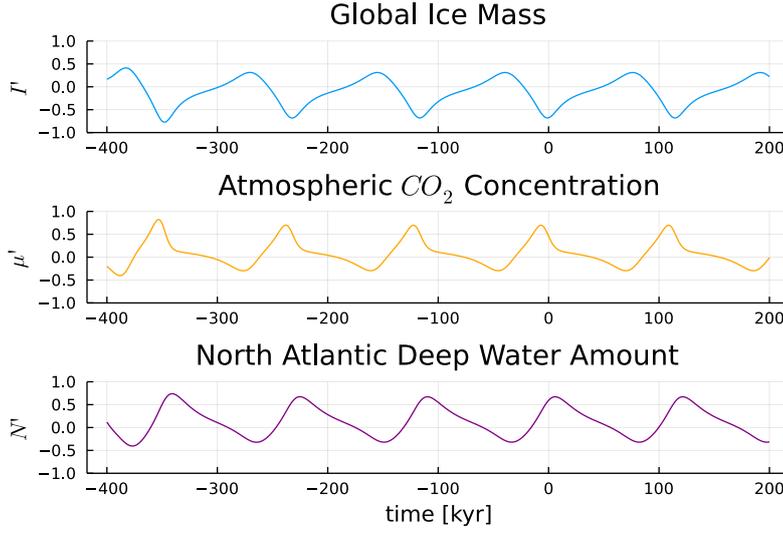


Figure 4.2: Evolution through time for the three variables.

### 4.3 Fixed Points

Then, the fixed points are searched. Those points are given when all the equations are cancelled,

$$\begin{aligned}\dot{X} &= -X - Y = 0 \\ \dot{Y} &= -pZ + rY + sZ^2 - Z^2Y = 0 \\ \dot{Z} &= -q(X + Z) = 0\end{aligned}$$

Solving the previous system, the following points are found:

$$\begin{aligned}\mathbf{x}_I &= \{0, 0, 0\}, \\ \mathbf{x}_{II,III} &= \left\{ \frac{1}{2}(-s \pm \sqrt{s^2 - 4(p-r)}), -X, -X \right\}.\end{aligned}$$

Thus, for the given parameter values:

$$\begin{aligned}\mathbf{x}_I &= \{0, 0, 0\}, \\ \mathbf{x}_{II} &= \{-0.15505, 0.15505, 0.15505\}, \\ \mathbf{x}_{III} &= \{-0.64495, 0.64495, 0.64495\}.\end{aligned}$$

As a following step, the stability of the fixed points is studied using the Jacobian, which for the given system is:

$$\mathbf{J} = \begin{bmatrix} -1 & -1 & 0 \\ 0 & r - Z^2 & -p + 2Z(s - Y) \\ -q & 0 & -q \end{bmatrix}$$

For the first fixed point,  $\mathbf{x}_I$ , the following eigenvalues are found:

$$\lambda_1 = -1.7575, \lambda_{2,3} = 0.1787 \pm 0.1906i.$$

While for the second fixed point,  $\mathbf{x}_{II}$ , it is found:

$$\lambda_1 = -1.6919, \lambda_2 = -0.1340, \lambda_3 = 0.4019.$$

Finally for the third fixed point,  $\mathbf{x}_{III}$ :

$$\lambda_1 = -1.7372, \lambda_{2,3} = -0.0394 \pm 0.4655i.$$

Therefore, the first and the second fixed points result to be unstable. Since for the first point, the real part of the complex conjugate eigenvalues is positive. While for the second point, three real eigenvalues are found, being one of them positive. However, for the third point, both the real eigenvalue and the real part of the complex conjugate eigenvalues are negative, presenting then a stable behavior.

The fact that the first and the second fixed points are unstable, means that any trajectory passing close to this points will not approach them, but will continue other trajectory. In particular, what happens is that those trajectories close enough to the first and second fixed points, will be attracted by the third fixed point, as this one is stable, or they will be dragged to the attractor. This behavior is explained in detail for each point.

### 4.3.1 Fixed Point $x_I$

As the first fixed point is unstable, determining the trajectory of the points around it is an important task. To do so, a sphere of points around  $x_I$  is evolved through time. The points were created equispacing the two angles of the spherical coordinates, that is, equispacing the polar angle between 0 and  $\pi$  radians,  $\theta = [0, \pi]$ , and the azimuth angle between 0 and  $2\pi$  radians,  $\varphi = [0, 2\pi]$ . Due to the particularities of the spherical coordinates, if  $N$  values of the polar angle equispaced in the previous described range are combined with  $N$  values of the azimuth angle equispaced in the correspondent range, instead of getting  $N^2$  unique points, they are obtained  $M$  unique points following the relation  $M = N(N - 1) - 2(N - 2)$ . The first obvious reason is that 0 and  $2\pi$  radians are the same angle, that is what causes  $N(N - 1)$  instead of  $N \cdot N$ . The second reason is that when the polar angle equals 0 or  $\pi$  radians, the azimuth angle is arbitrary, since  $x = \sin \theta \cos \varphi$  and  $y = \sin \theta \sin \varphi$ , thus, the coordinates  $x$  and  $y$  will be 0 for all the different values of the azimuth angle. Therefore, for  $N - 1$  points that are generated when  $\theta = 0$ , only one unique point is generated, same for  $\theta = \pi$ , which only generates other unique point. So, taking into account these particularities,  $M$  points are generated for different values of the radius. In particular, two different values of the radius were considered enough to show the behavior of the several fixed points. To exemplify, in Figure 4.3 a set

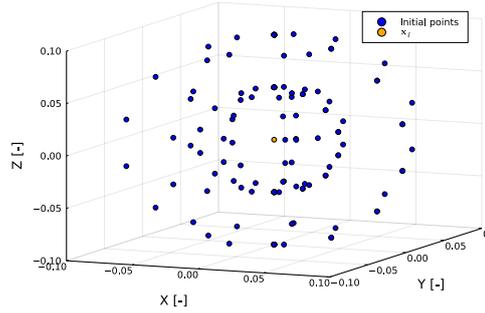


Figure 4.3: Set of initial points around first fixed point  $x_I$ .

of  $M$  initial points at radius 0.1 along with other  $M$  initial points at radius 0.05 are depicted around the first fixed point. The number  $M = 44$  comes from the previous relation when  $N = 8$  is used.

In Figure 4.4, a set of 28 points contained in a sphere of radius 0.1 are evolved till  $t = 80$ , and as it is showed, their trajectories are split in two different destinations. In particular, these two destinations happen to be the attractor and the stable fixed point  $x_{III}$ , as it could be appreciated. In Figure 4.4, 14 points at a distance of 0.1 from  $x_I$  and 14 points at a distance of 0.05 are evolved. However, even if this figure provides the big picture of the behavior, it does not show in detail which trajectories are dragged to the attractor and which ones are lead to the stable fixed point.

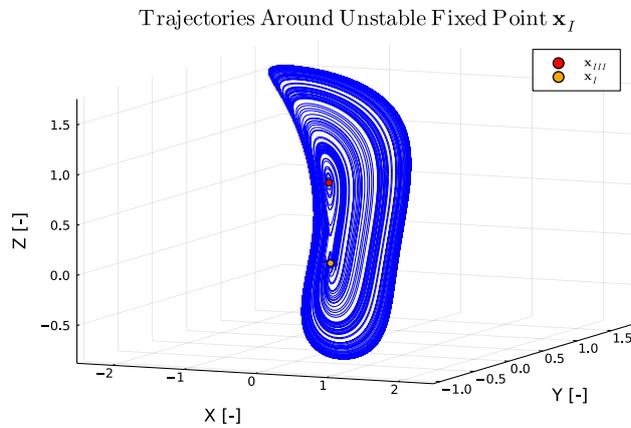


Figure 4.4: Trajectories of points around the unstable fixed point  $x_I$ .

Thus, in Figure 4.5 two figures were obtained focusing first in the unstable fixed point  $x_I$ , Figure 4.5a, and second in the stable fixed point  $x_{III}$ , Figure 4.5b.

In the figure around  $x_I$ , Figure 4.5a, it can be seen how the trajectories are split in two, in opposite directions. One will reach the attractor, while the other one will be drawn to the stable fixed point,  $x_{III}$ . In the figure, 44 points at 0.05 distance along with 44 points at radius 0.025 were evolved. The radius to obtain the figure were decreased in order to show a clear picture of the behaviour.

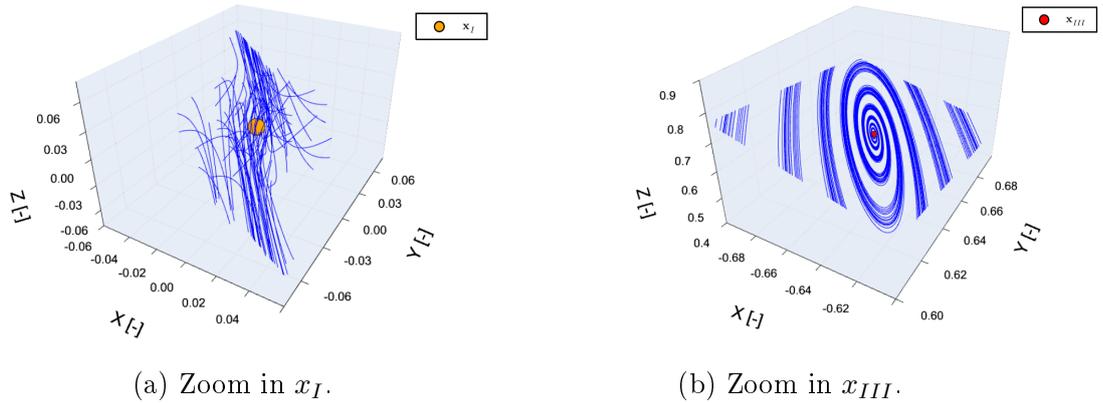


Figure 4.5: Zoom of trajectories around unstable fixed point  $x_I$ .

In the figure around  $x_{III}$ , Figure 4.5b, it can be seen that the trajectories are contained in a plane with the point  $x_{III}$  belonging to it. The trajectories "orbit" the point while they are getting closer, forming a spiral. In this case, the trajectories are evolved till  $t = 150$ , since reaching the fixed point takes more time than arriving to the attractor.

### 4.3.2 Fixed Point $x_{II}$

The same steps are followed to study the behaviour around the second fixed point, as this one is unstable as the first one. Therefore, in Figure 4.6, the picture is showed, where again, some points are dragged to the attractor while others arrive to the stable fixed point,  $x_{III}$ . In this case, 44 points were created at 0.1 distance from the unstable fixed point  $x_{II}$  and other 44 points at a radius of 0.05. The trajectories were evolved during  $t = 80$  to obtain the figure. Comparing to the first fixed point, the trajectories here follow a more synchronized and faster path to the attractor. While the path to the stable fixed point remains very similar.

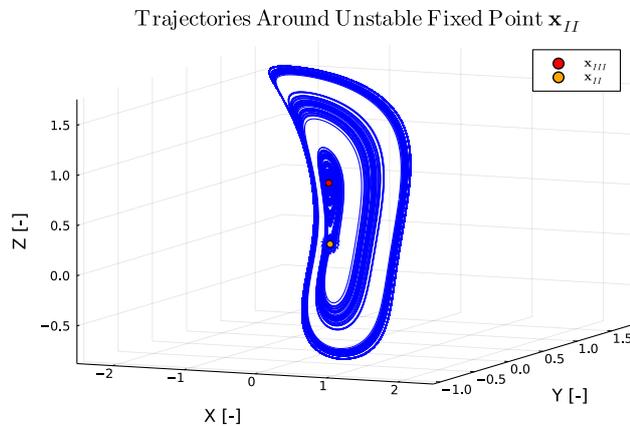


Figure 4.6: Trajectories of points around the unstable fixed point  $x_{II}$ .

Thus, focusing again in the unstable fixed point treated in this section and the stable fixed point, the Figure 4.7 is obtained. In Figure 4.7a, an amount of 44 points were placed at 0.05 distance from the fixed point and 44 points were placed at 0.025. In this case, again, the distance from the fixed point is reduced to get to capture a clearer figure. Same set is used in Figure 4.7b, but in this case the trajectories showed were evolved till  $t = 150$ , as it is the time they needed to reach the final destination, i.e. the stable fixed point  $x_{III}$ .

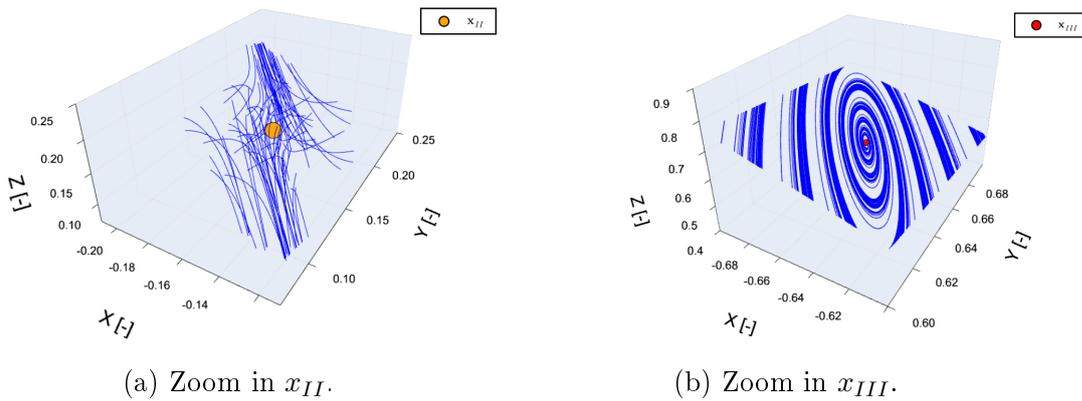


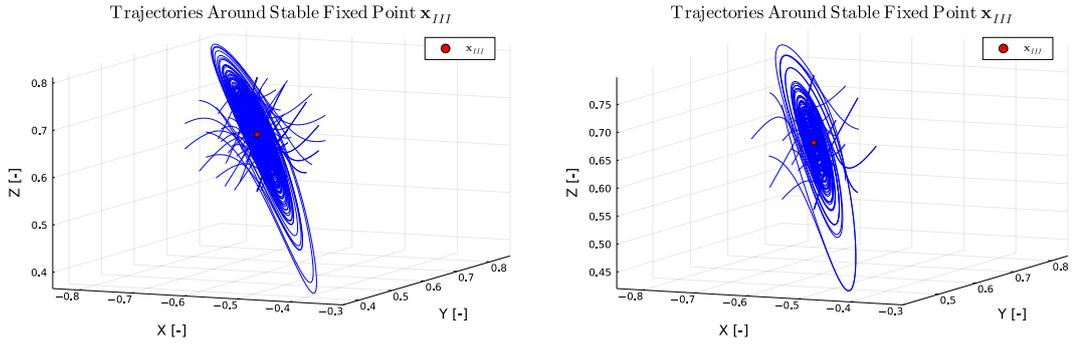
Figure 4.7: Zoom of trajectories around unstable fixed point  $x_{II}$ .

### 4.3.3 Fixed Point $x_{III}$

As aforementioned, the fixed point  $x_{III}$  is stable, meaning that trajectories beginning in its surroundings will be attracted to the point itself. A study of its sensitivity has been conducted, concluding that points contained in a sphere of radius  $\leq 0.12$  will always be attracted to the fixed point. For bigger radius the trajectories start to escape the attraction to the point and instead they go to the attractor.

In the Figure 4.8, the explained behaviour is illustrated. In Figure 4.8a, 44 points at a distance of 0.12 from  $x_{III}$  and 44 points at a distance of 0.06 are evolved through time. Alternatively, in Figure 4.8b, less points were evolved with the purpose of presenting a clearer figure where it could be appreciated the trajectory of the points. In particular, an amount of 28 points were evolved. In a similar way as the previous simulation (Figure 4.8a), half the points (14) started at a radius of 0.12 and the other half started at a distance of 0.06 from the point.

Same behaviour is found in both figures, all the points are lead to a particular plane in which they continue their trajectory "orbiting" the point. Diminishing the distance from the point little by little, that is, following a spiral.



(a) Trajectories of 88 points contained in a sphere of radius = 1.2. (b) Trajectories of 28 points contained in a sphere of radius = 1.2.

Figure 4.8: Trajectories of points around the stable fixed point  $x_{III}$ .

## 4.4 Attractor

Another property of the dynamical system, along with the fixed points, which can tell a valuable information about the dynamics of the system is the attractor. For the case studied here, there is one attractor that attracts almost all the possible trajectories inside the phase space, except for some trajectories that pass close the three fixed points, which end in the stable one. Once the trajectory is captured by the attractor, the particle will follow a periodic trajectory along the attractor, thus, it could be also called limit cycle. This particularity explains the oscillatory behavior that Saltzman was searching for.

In Figure 4.9, seven random initial points are evolved and their trajectories depicted. Thus, the shape of the attractor is easily captured computing the trajectory of a random initial point. It is showed in Figure 4.10.

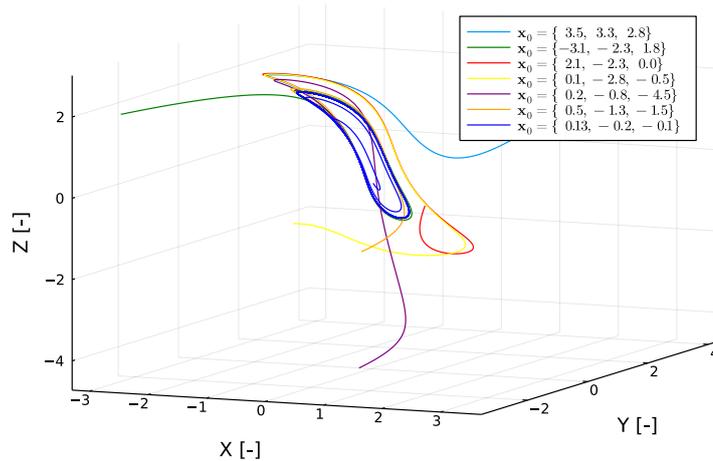


Figure 4.9: Example of trajectories being dragged to the attractor.

Since, the issue of the domain will be treated later, the knowledge of the limits

of the attractor is important, in particular they are:

$$\begin{cases} X \in (-2.06, 0.943) \\ Y \in (-1.17, 2.77) \\ Z \in (-0.801, 1.68) \end{cases} \quad (4.12)$$

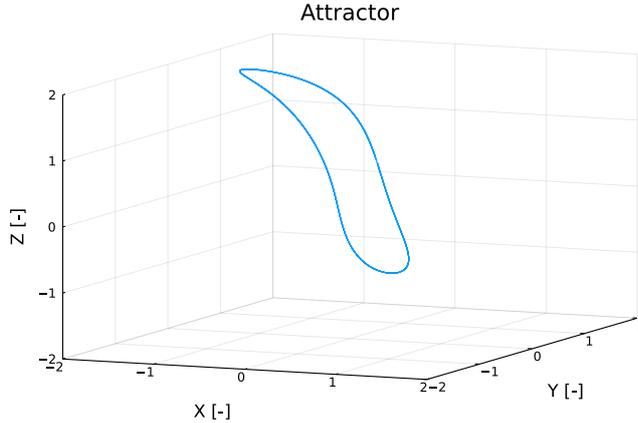


Figure 4.10: Attractor found for the dynamical system presented by Saltzman and Maasch.

## 4.5 Phase Space

Mathematically, the phase space of the system goes from  $-\infty$  to  $\infty$  for every dimension, i.e.  $X \in (-\infty, \infty)$ ,  $Y \in (-\infty, \infty)$ ,  $Z \in (-\infty, \infty)$ . Yet, if the physics of the system are considered, points far from  $(0, 0, 0)$  would not be included, as they lack the physical meaning. In fact, as pointed out in Saltzman and Maasch (1988), “*When the system is very far from this unstable equilibrium large stabilizing restorative forces come into play, constraining the system to execute oscillatory variation about the unstable equilibrium.*” Where “*unstable equilibrium*” refers to the fixed point at the origin of coordinates. So, in practice, the phase space could be reduced to the surroundings of the attractor described before, which corresponds to the “*oscillatory variation*” mentioned in the article.

In Figure 4.11, the phase portrait of the dynamical system is presented. The same figure is presented with two different orientations, so that the particular paths that the trajectories follow could be appreciated. To create the picture, a mesh of initial equispaced points forming the surfaces of a cube which length of the edges is 6 and that is centered in the origin, are evolved until they reach the attractor. In Figure 4.11, the trajectories of these points are depicted in dark blue. In addition to that, trajectories with initial points around the three fixed points and in general, inside the attractor, are evolved. However, what happens inside the attractor is not well appreciated in the figure.

In order to clearly present the phase portrait inside the attractor, the trajectories coming from this cube centered in the origin, are depicted from a time  $t > 0$ , which

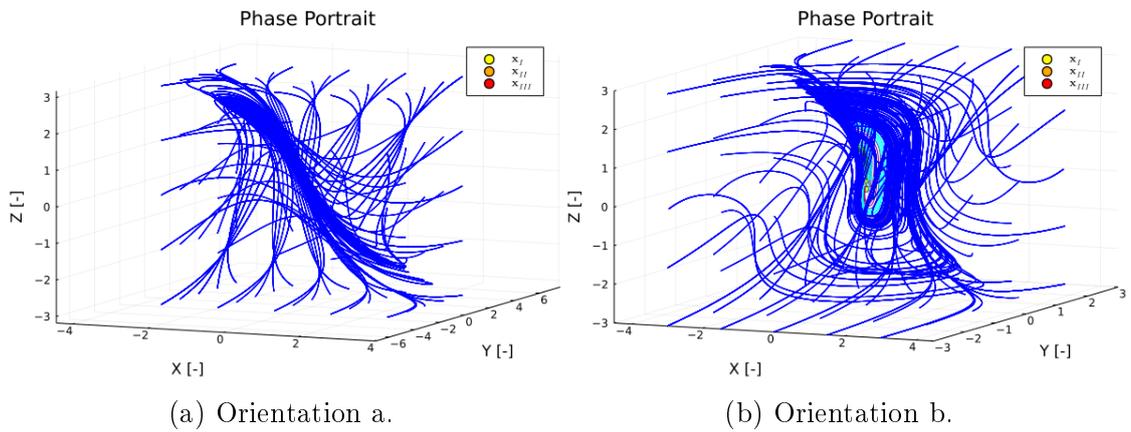


Figure 4.11: Phase Portrait.

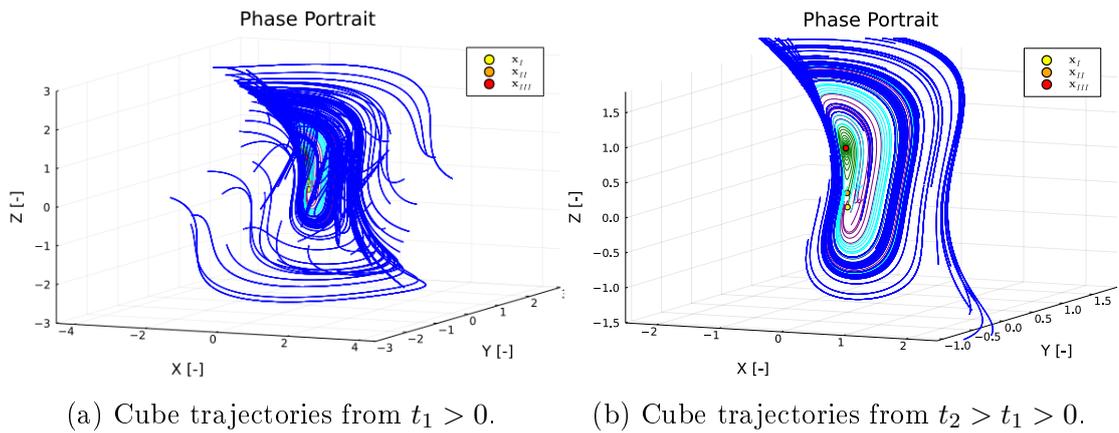


Figure 4.12: Phase Portrait.

deletes the first part of their paths making it possible to present a more clear picture. The result is showed in Figure 4.12.

Lastly, a phase portrait where only trajectories coming from the inside of the attractor is showed in Figure 4.13. Here, it can be appreciated more in detail the behavior around the fixed points mentioned before. In green, trajectories coming from the unstable fixed points  $x_I$  and  $x_{II}$  and reaching the stable fixed point  $x_{III}$  are depicted. As showed in the correspondent sections before, not all the points around the unstable fixed points are attracted to the stable one, but a part of them are dragged to the attractor. Those cases are depicted in cyan when coming from  $x_{II}$  or in purple when coming from  $x_I$ .

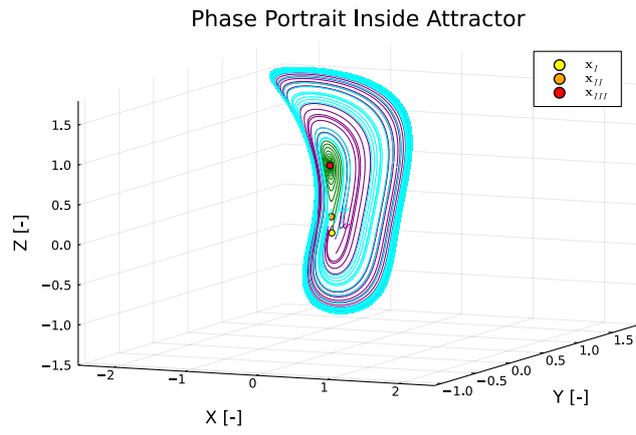


Figure 4.13: Phase Portrait inside the attractor.

# Chapter 5

## Probability Density Function

As aforementioned, the goal is to compute the Equation 3.12. To compute the ensemble averages that are present in the equation, the initial probability density function,  $f_0$ , is needed. This can be given by knowledge of the problem at hand, and the initial state of the system of interest, or it could in any case be determined by convenience. Like in any time evolution, the initial condition must be known. If it is not suggested by the problem itself, like in the case of particle systems, where it is usually an equilibrium distribution, the simplest compatible with the theory is the best choice.

In the case of particle systems, as in Caruso et al. (2020), one may take the Gibb's canonical distribution, that is,  $f_0(\mathbf{x}) = \frac{e^{-\beta H(\mathbf{x})}}{Z_\beta}$ , as the initial distribution like mentioned before. Nevertheless, for generic dynamical systems, the case treated here, since the Dissipation Function  $\Omega$ , does not lead to thermodynamic-like relations with physical meaning, the initial distribution taken does not need that physical meaning.

The most trivial case would be if the system preserves the volume. In fact, the initial distribution can be chosen uniform, a constant value for the whole phase space,  $f_0 = \frac{1}{V}$ , being  $V$  the volume of the phase space. Which can be easily deduced knowing that it must be normalized in the phase space.

However, the dynamical system discussed here does not preserve its volume, but shrinks. Thus, several strategies to obtain a suitable initial probability density function are explained through this chapter.

### 5.1 Uniform Distribution

Just like it can be safely done with systems that preserve the volume, a constant value could be chosen as the initial probability density function (pdf). Therefore the initial pdf would be  $f_0 = \frac{1}{V}$  in the chosen domain, where  $V$  is the correspondent volume of the given domain, and  $f_0 = 0$  out of that domain. This can be done, thanks to the fact that it is known that a volume around the origin contains trajectories that never leave the volume. As they are all attracted to the attractor or the stable

fixed point. The issue faced here is that unlike the case of dynamics preserving the volume, it cannot be said that no trajectory is going to enter the chosen the volume if it is chosen arbitrarily. Which translates in obtaining an improper initial pdf, as it will not be normalized.

In order to solve the issue, two options are presented. The first one would be finding a volume where in addition to assuring that the trajectories do not leave it, there are no trajectories entering it. Thus, the volume of that particular domain is preserved and no more special treatment to this domain has to be considered. While the second option considers an arbitrary volume, only centered in the origin and containing the attractor and stable fixed point so that the trajectories do not escape. But, unlike the first option, the domain requires some more special treatment. Both options are explained hereafter.

### 5.1.1 Domain Preserving its Volume

The condition which assures that trajectories are not entering nor escaping a given volume is that the derivatives along its boundaries are equal to zero. Hence, using the equations of the dynamics of the system and searching the surfaces where they are null, these boundaries of the searched volume,  $V^*$ , could be obtained. Once the domain is found, the initial pdf would be simply:

$$f_0(\mathbf{x}) = \begin{cases} \frac{1}{V^*} & \mathbf{x} \in V^* \\ 0 & \mathbf{x} \notin V^* \end{cases} \quad (5.1)$$

In the case treated here, the first bound for that searched volume would be, using Equation 6.6:

$$\dot{X} = -X - Y = 0 \quad \rightarrow \quad X = -Y$$

However, if a trajectory comes towards this first boundary with a component in the  $Y$  direction, this plane where the velocity along the  $X$  direction equals 0, could be crossed. Hence, it would not keep the trajectories out of the volume delimited by itself and the other boundaries.

The second simple boundary, is obtained from the Equation 6.8:

$$\dot{Z} = -q(X + Z) = 0 \quad \rightarrow \quad Z = -X$$

Finding the same result, a trajectory with  $Z$  component in the velocity would be able to enter the volume through this boundary.

Finally the third one, obtained using Equation 6.7 is as follows:

$$\dot{Y} = -pZ + rY + (s - Y)Z^2 = 0 \quad \rightarrow \quad Z = \frac{p \pm \sqrt{p^2 - 4(s - Y)rY}}{2(s - Y)}$$

The third and last condition, provides a more complex surface, which was not studied, since, analyzing the results obtained, it is clear that the searched volume

where the trajectories do not enter nor escape, does not exist for the given dynamics. And therefore, other strategies will be chosen for the definition of the initial probability density function.

### 5.1.2 Pseudo-arbitrary Domain

For this apparently simpler strategy of obtaining the initial probability density function, the domain is chosen in a more arbitrary way, i.e. only assuring that the contained trajectories will not leave it. That condition, as aforementioned, is easy to fulfill taking into account that all the trajectories are dragged to the attractor or attracted to the stable fixed point. That way, there exist infinite number of volumes that could be chosen, with the only condition that they contain both the attractor and the stable fixed point. Hence, the simplest volume that came to mind was a cube centered in the origin whose length of the edges is big enough to contain the attractor, bigger than the domain which contains the attractor, presented before Equation 4.12.

As unknown trajectories will enter the domain while the time grows, there is the risk that the probability density function is not normalized anymore. Thus, to solve the problem what it is actually done is evolving the domain backwards in order to obtain the initial volume which originates the chosen cube. And so, the chosen cube,  $V_t$ , is the domain of the probability density function at a future time, while the domain of  $f_0$ ,  $V_0$ , is the obtained domain evolving the cube backwards in time. Similar to Equation 5.1, the expression of the initial probability density in this case would be:

$$f_0(\mathbf{x}) = \begin{cases} \frac{1}{V_0} & \mathbf{x} \in V_0 \\ 0 & \mathbf{x} \notin V_0 \end{cases} \quad (5.2)$$

This particular treatment of the  $f_0$  domain, is conducted in the following example that can be found in Caruso et al. (2020), presented here in order to exemplify it in a more clear way.

The dynamical system considered is the following

$$\dot{x} = -x; \quad x \in \mathbb{R}$$

so the initial density is

$$f_0(x) = \begin{cases} 1/2 & \text{if } x \in [-1, 1] \\ 0 & \text{else} \end{cases}$$

Then, knowing that the initial condition  $x$  evolves like  $\mathcal{S}^t x = e^{-t}x$ ,  $f_t$  is given by

$$f_t(x) = \frac{1}{2} e^t \chi_{[-e^{-t}, e^{-t}]}(x), \quad \text{with } \chi_{[-e^{-t}, e^{-t}]}(x) = \begin{cases} 1 & \text{if } x \in [-e^{-t}, e^{-t}] \\ 0 & \text{if } x \notin [-e^{-t}, e^{-t}] \end{cases}$$

Therefore in the example,  $V_0 = [-1, 1]$  and  $V_t = [-e^{-t}, e^{-t}]$ . Obviously in the case studied here, the dynamics are much more complex and the search for the

resulting volume is not trivial. In fact, the big problem that appears is that evolving the dynamics backwards begins to be unstable when it reaches a certain value of  $t$ . Because of this particularity, the option presented here does not seem to be the optimal one and so, it was discarded for the computing of the solution.

## 5.2 3D Gaussian Distribution

Finally, the last strategy, and the chosen one, is using a 3 dimensions Gaussian distribution as initial probability density function. That way, the domain of the function goes from  $-\infty$  to  $\infty$  in every dimension, so the mathematical phase space of the system treated here, and there is no need to worry about the change in the volume. Choosing a Gaussian with standard deviation equal to 3.5 would be enough to properly contain the attractor and the stable fixed point, as it could be deduced from Equation 4.12.

Following the Equation 3.7:

$$f_{s+t}(\mathbf{x}) = \exp\{\Omega_{-t,0}^{f_s}(\mathbf{x})\} f_s(\mathbf{x})$$

the function  $f_t$  could be obtain.

Even though the problem of computing the evolution of the probability distribution could be solved, as discussed here, it remains that probabilities are immaterial, although very useful. What matters is the evolution of observable quantities, i.e. Equation 3.12. The theory allows one to bypass the explicit calculation of the probability density at a given time  $t > 0$ , and to directly compute the behavior of measurable quantities. The mentioned computation is explained in the following chapter, chapter 6.

# Chapter 6

## Numerical Computation of the Exact Response Theory

### 6.1 Perturbed Vector Field

As aforementioned, the Exact Response Theory presented here allows one to compute the averaged evolution of a measurable observable after a perturbation in the system. Moreover, the theory is able to deal with perturbations which are not necessarily small, contrary to other theories. Thus, given the unperturbed dynamical system which follows the equation

$$\dot{\mathbf{x}} = \mathbf{V}_0(\mathbf{x}) \quad (6.1)$$

the perturbed dynamics can be written as the following

$$\dot{\mathbf{x}} = \mathbf{V}_0(\mathbf{x}) + \epsilon \mathbf{V}_p(\mathbf{x}) \quad (6.2)$$

being  $\epsilon$  the intensity of the perturbation, which is not restricted to small numbers.

For the application of the theory to the system provided by Saltzman and Maasch (1988) conducted in this work, it was decided to perturb the parameters of the system. In particular, the parameter  $r$  that appears in the second equation of the system, that way, the perturbed dynamics are

$$\dot{X} = -X - Y \quad (6.3)$$

$$\dot{Y} = -pZ + rY + (\epsilon Y) + sZ^2 - Z^2Y \quad (6.4)$$

$$\dot{Z} = -q(X + Z) \quad (6.5)$$

or written in a compact way

$$\dot{X} = -X - Y \quad (6.6)$$

$$\dot{Y} = -pZ + \tilde{r}Y + sZ^2 - Z^2Y \quad (6.7)$$

$$\dot{Z} = -q(X + Z) \quad (6.8)$$

where  $\tilde{r} = r + \epsilon$ . The value for the intensity of the perturbation was defined as  $\epsilon = 0.1$ . Being the value of  $r = 0.8$ , the intensity of the perturbation represents a 12.5% of the unperturbed value of the parameter  $r$ . Hence, the perturbation is not considered a small perturbation.

For the physics of the system, the perturbation of the parameter that multiplies the variable  $Y$ , i.e. atmospheric  $CO_2$  concentration, would represent a possible change in the way the amount of  $CO_2$  concentration affects directly its increase or decrease. However, since  $r = b_1/a_0 = (r_1\beta + r_2e_\mu + r_5a_1 - r_4)/(s_4 - s_1\alpha - s_3e_I)$ , checking in Equation 4.1,  $r_1, r_2, r_4, r_5$  multiply the mean oceanic surface temperature, the permanent sea ice extent, the  $CO_2$  itself and the global ice mass, respectively. Hence, it could also represent a change in the way these former variables relate to the  $CO_2$  concentration. For example, a change in the solubility of  $CO_2$  in the ocean, as the oceanic temperature affects the atmospheric  $CO_2$  by increasing the solubility and uptake when colder waters.

## 6.2 Computing the Evolution of the Observable $X$

As stated before, the goal is to compute the averaged evolution of a measurable observable, therefore, Equation 3.12 must be computed. In the present work, looking for simplicity, one of the variables of the dynamical system is chosen as the studied observable. In particular, the variable  $X$ , i.e. the global ice mass ( $I$  before the normalization), is chosen.

Therefore, the first summing term of the Equation 3.12 becomes:

$$\langle \mathcal{O} \rangle_{f_0} = \int_{\mathcal{M}} \mathcal{O}(\mathbf{x}) f_0(\mathbf{x}) d\mathbf{x} = \int_{\mathcal{M}} X f_0(\mathbf{x}) d\mathbf{x}$$

where  $\mathcal{M}$  is the found domain of  $f_0$ , which is taken as the phase space of the dynamical system (which extends to infinity), since the chosen initial probability density function is a 3D gaussian distribution, as explained previously in chapter 5.

While the second summing term of the equation can be written as:

$$\int_0^t \left[ \int_{\mathcal{M}} \{(\mathcal{O} \circ \mathcal{S}^s) \cdot \Omega^{f_0}\} f_0(\mathbf{x}) d\mathbf{x} \right] ds = \int_0^t \left[ \int_{\mathcal{M}} \{(\mathcal{O}(\mathbf{x}(s))) \cdot \Omega^{f_0}(\mathbf{x})\} f_0(\mathbf{x}) d\mathbf{x} \right] ds.$$

Thus, in order to compute the integral, the expression of the Dissipation Function,  $\Omega^{f_0}$ , must be considered. Using the following relation

$$\Omega^{f_0}(\mathbf{x}) := -[\Lambda(\mathbf{x}) + \mathbf{V}(\mathbf{x}) \cdot \nabla \ln f_0(\mathbf{x})], \quad (6.9)$$

where,  $\Lambda(\mathbf{x}) = \nabla \cdot \mathbf{V}(\mathbf{x})$ , being  $\mathbf{V}(\mathbf{x})$  the perturbed system.

The expression of  $\Lambda$  for the perturbed field is

$$\Lambda(\mathbf{x}) = \frac{\partial V_X}{\partial X} + \frac{\partial V_Y}{\partial Y} + \frac{\partial V_Z}{\partial Z} = (-1) + (\tilde{r} - Z^2) + (-q)$$

### 6.2.1 Code Implementation

Putting the theory into practice, the equations Equation 6.9 and section 6.2 and the functions presented before were implemented in Julia Programming Language. The developed code may be consulted in Appendix A.

In order to compute the integrals in the phase space, an ensemble of random points was created. In particular, the points were created following an uniform random distribution along  $(-3, 3)$  for every dimension. The initial probability density function  $f_0$  was chosen a 3D Gaussian distribution, as explained in chapter 5, with mean  $\mu = (0, 0, 0)$  and standard deviation  $\sigma = 3.5$  for every dimension.

Then, through a loop, the Equation 6.9 is computed for every point in the created set, and thus the integral value is obtained.

While for the second integral, section 6.2, it is needed a loop over the time, as well as the loop over the phase space. Since a simple numerical integration was conducted. In particular, the chosen step of the time integral was  $\Delta t = 0.1$ . And the solution was computed with final time  $t_f = 4$ . This second integration is more complex, as the observable needed to be evolved with every step, while  $\Omega^{f_0}$  and  $f_0$  were evaluated in the initial points.

### 6.2.2 Results

Given the computational costs of performing simulations with large sets of points, the result was obtained computing the mean of several simulations with different number of initial points. More in detail, 10 simulations of 1000000 initial points, 7 simulations with 15625000 points, 3 with 64000000 initial points and 1 simulation of 125000000 initial points, was the combination. In Figure 6.1, the mean and the standard deviation of the ensemble of simulations is presented.

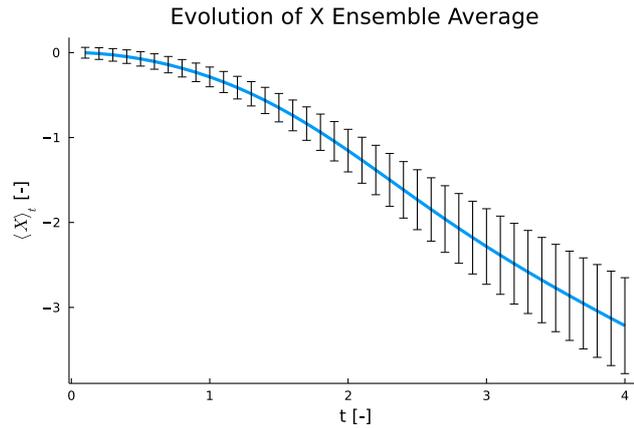


Figure 6.1: Evolution of the ensemble average of the observable  $X$  after the perturbation in the parameter  $r$ .

However, even if important, no data were available to confront the results and

validate the simulations. Therefore, the results obtained here are not presented as prove of the performance of the method, but as a first approximation to the procedure. And consequently, it is emphasized that future studies should search for a way to validate the results.

# Chapter 7

## Conclusions

The climate system is forced, dissipative, nonlinear, chaotic, and out of thermodynamic equilibrium. Along with the inherent difficulties that this type of systems faces to model and predict the response to perturbations, the climate system presents some particularities such as, the presence of well-defined subsystems, the continuous variation of the forcings changing the atmospheric composition, the lack of scale separation, the lack of complete and useful observations, the fact that only one realization of the processes involved in climate evolution is available, which add more obstacles to the task. Principally, it is hard to separate the climate system's response to different forcings from its natural variability. Therefore, the approaches to the problem need to take special care to these features.

The interest in the matter is not purely scientific, in fact, the anthropogenic contribution to the climate evolution is currently a central topic, as the severity of its consequences is driving a considerable amount of efforts trying to mitigate this climate change. The Intergovernmental Panel on Climate Change (IPCC) releases its assessment reports based on the best science at hand, where they summarize the progress in observing, modeling, understanding and predicting the evolution of the climate system. The methods used in the analysis of climate response were reviewed, finding that the new perspectives in the area of dynamical systems could provide powerful and useful tools.

Even if it arises as a new method which would reduce considerably the computational cost of simulating climate response to perturbations, the approach developed by Lucarini (Lucarini et al. (2017)) in this context of dynamical systems, faces some limitations that were critically reviewed here. In particular, it is based on Ruelle's response theory (Ruelle (1998)), which is a linear response theory, thus, restricts the study to small perturbations. A second limitation comes from the assumption of the climate system being an Axiom A system, since Ruelle's theory is developed for that kind of systems. Finally, they are forced to obtain the operator they need through a set of experiments since putting Ruelle's theory into practice for complex models like climate system is far from trivial.

The Exact Response Theory incorporates the Dissipation Function,  $\Omega$ , as its basis. It is expected to predict the response of a system with many degrees of freedom, even if arbitrarily large perturbations and modification of states are present. In the context of particle systems, the Dissipation Function is associated with the entropy production rate. Also, the Dissipation Function is able to capture meaningful behaviors of the system, something that the classical linear response theory could never achieve.

Recently, the Exact Response Theory has been successfully applied to systems such as the Kuramoto Oscillators reviewed here, and applying it to the problem of obtaining the climate system response to perturbations could add one more method to the new tendency of incorporating dynamical systems machinery. Providing the advantages of being applicable, without making non rigorous assumptions and obtaining a response to non necessarily small perturbations. The climate model proposed in Saltzman and Maasch (1988), describing the evolution of climate during the Pleistocene ice ages, was studied and used as the system to put into practice the theory. The oscillation in the variables was found to be caused by the presence of an attractor.

The simulations performed here, obtaining the ensemble average of the observable accounting for the global ice mass after a perturbation, however, were not confronted since no appropriate data was available to do so. Hence, the result of the simulations is presented as a first approach to the application of the exact response theory to the complex climate response problem, and it is vital to find a way to validate the outputs in the future studies.

Moreover, a more realistic climate model could be proposed to be used for the validation of the suitability of the theory. For example, even a model like PlaSim, the one used by Lucarini, which does not include as many degrees of freedom as the models currently used in the IPCC reports, but that includes a considerable number, seems to be interesting to consider.

# Appendix A

## Julia code

```
##-----Main functions-----##  
  
# Dynamical system  
function carbon(du,vec,p,t)  
    p,q,r,s = 0.9, 1.2, 0.8, 0.8  
    x,y,z = vec  
    du[1] = -x -y  
    du[2] = -p*z + r*y + s*z^2 - z^2*y  
    du[3] = -q*(x+z)  
end  
  
# V  
function V(x_vec)  
    x,y,z = x_vec  
    p,q,r,s = 0.9, 1.2, 0.8, 0.8  
    dx = -x -y  
    dy = -p*z + r*y + s*z^2 - z^2*y  
    dz = -q*(x+z)  
    dxyz = [dx,dy,dz]  
    return dxyz::Vector{Float64}  
end  
  
# Lambda of carbon cycle  
function Lambda_carbon(x_vec::Vector{Float64})  
    X,Y,Z = x_vec  
    p,q,r,s = 0.9, 1.2, 0.8, 0.8  
    Lambda = -1 + (r-Z^2) - q  
    return Lambda  
end
```

```

# Grad_lnf
function Grad_lnf(f::Function, x_vec::Vector{Float64})
    x,y,z = x_vec
    h = 0.005
    if f(x,y,z) == 0.0
        Grad_lnf_val = [0.0,0.0,0.0]
    else
        df_dx = 1/(2*h)*(f(x+h,y,z)-f(x-h,y,z))
        df_dy = 1/(2*h)*(f(x,y+h,z)-f(x,y-h,z))
        df_dz = 1/(2*h)*(f(x,y,z+h)-f(x,y,z-h))
        Grad_lnf_val = 1/(f(x,y,z))*[df_dx, df_dy, df_dz]
    end
    return Grad_lnf_val::Vector{Float64}
end

# Initial pdf
function f_0(x,y,z)
    IsoNormal = MvNormal([0.0,0.0,0.0], 3.5*Matrix(I,3,3) )
    return pdf(IsoNormal,[x,y,z])
end

# Computing of Omega^f_0 generic
function Omega(x_vec::Vector{Float64}, Lambda::Function, V::Function,
               Grad_lnf::Function, f::Function)
    V_vec = V(x_vec)
    Grad_lnf_vec = Grad_lnf(f,x_vec)
    prod = 0.0
    for i in 1:3
        prod += V_vec[i]*Grad_lnf_vec[i]
    end
    Omega = - (Lambda(x_vec) + prod)
    return Omega
end

# Observable function
function Obs(x_vec::Vector{Float64})
    return x_vec[1]
end

```

```
## Perturbed system functions
```

```
epsilon = 0.1
```

```
# Perturbed vector field
```

```
function carbon_per(du,vec,p,t)
```

```
    p,q,r,s = 0.9, 1.2, 0.8, 0.8
```

```
    x,y,z = vec
```

```
    du[1] = -x -y
```

```
    du[2] = -p*z + r*y + s*z^2 - z^2*y + epsilon*y
```

```
    du[3] = -q*(x+z)
```

```
end
```

```
# Perturbed V
```

```
function V_per(x_vec)
```

```
    x,y,z = x_vec
```

```
    p,q,r,s = 0.9, 1.2, 0.8, 0.8
```

```
    dx = -x -y
```

```
    dy = -p*z + r*y + s*z^2 - z^2*y + epsilon*y
```

```
    dz = -q*(x+z)
```

```
    dxyz = [dx,dy,dz]
```

```
    return dxyz::Vector{Float64}
```

```
end
```

```
# Perturbed Lambda
```

```
function Lambda_per(x_vec::Vector{Float64})
```

```
    X,Y,Z = x_vec
```

```
    p,q,r,s = 0.9, 1.2, 0.8, 0.8
```

```
    Lambda = -1 + (r-Z^2 + epsilon) - q
```

```
    return Lambda
```

```
end
```

```
##----- Compute O_f0 (first integral of the equation) -----##
```

```
# Creating initial domain
```

```
a = -3.0
```

```
b = 3.0
```

```
n_points = 250
```

```
uni_dis = Uniform(a,b)
```

```
x_0 = rand(uni_dis, n_points)
```

```
y_0 = rand(uni_dis, n_points)
```

```
z_0 = rand(uni_dis, n_points)
```

```
O_f0 = 0.0
```

```
O_f01 = 0.0
```

```

Delta_x = ((b-a)/n_points)^3
for i in 1:length(x_0)
    for j in 1:length(y_0)
        for k in 1:length(z_0)
            global O_f0 += f_0(x_0[i], y_0[j], z_0[k])*
                Obs([x_0[i], y_0[j], z_0[k]])*Delta_x
        end
    end
end

##----- Compute second integral of the equation -----##

ind_t = 40
Second_int = zeros(ind_t)
Second_int_int = zeros(ind_t)
Second_int_int_0 = 0.0
t = zeros(ind_t)

# Auxiliary vector to store x values
x_last_t = zeros((n_points)^3,3)
p = 1
for i in 1:n_points
    for j in 1:n_points
        for k in 1:n_points
            x_last_t[p,:] = [x_0[i], y_0[j], z_0[k]]
            global p += 1
        end
    end
end

# Auxiliary function for 2nd integral
function fun(x_vec_0::Vector{Float64}, x_S::Vector{Float64},t)
    #Initial
    x_init = x_vec_0[1]
    y_init = x_vec_0[2]
    z_init = x_vec_0[3]
    fun = Obs(x_S)*Omega(x_vec_0, Lambda_per, V_per, Grad_lnf, f_0)
        *f_0(x_init,y_init, z_init)
    return fun
end

t0 = 0.0

```

```

Delta_t = 0.1
for tau in 1:ind_t
    q = 1
    t[tau] = (tau)*Delta_t
    for i in 1:length(x_0)
        for j in 1:length(y_0)
            for k in 1:length(z_0)
                t_span = (t0, t[tau])
                vec_init = [x_last_t[q,1], x_last_t[q,2],
                           x_last_t[q,3]]
                prob = ODEProblem(carbon_per, vec_init, t_span)
                sol = solve(prob, Tsit5(), reltol = 1e-8,
                           abstol=1e-8)

                x = sol[1,end]
                y = sol[2,end]
                z = sol[3,end]

                Second_int[tau] += fun([x_0[i], y_0[j], z_0[k]],
                                       [x,y,z],t[tau])*Delta_x
                x_last_t[q,:] = [x,y,z]
                q += 1
            end
        end
    end
end

Second_int_int[tau] = Second_int_int_0 +
                    Second_int[tau]*Delta_t
global Second_int_int_0 = Second_int_int[tau]
global t0 = t[tau]
end

O_t = O_f0 + Second_int_int[end]
O_tevol = O_f0*ones(ind_t) + Second_int_int

```



# Bibliography

- Amadori, D., Colangeli, M., Correa, A., and Rondoni, L. (2022). Exact response theory and kuramoto dynamics. *Physica D: Nonlinear Phenomena*, 429:133076.
- Barnola, J. M., Raynaud, D., Korotkevich, Y. S., and Lorius, C. (1987). Vostok ice core provides 160,000-year record of atmospheric co<sub>2</sub>. *Nature*, 329:408–414.
- Caruso, S., Giberti, C., and Rondoni, L. (2020). Dissipation function: Nonequilibrium physics and dynamical systems. *Entropy 2020, Vol. 22, Page 835*, 22:835.
- Curry, W. B. and Crowley, T. J. (1987). The  $\delta^{13}\text{C}$  of equatorial atlantic surface waters: Implications for ice age pco<sub>2</sub> levels. *Paleoceanography*, 2:489–517.
- Evans, D. J. and Searles, D. J. (2002). The fluctuation theorem. *undefined*, 51:1529–1585.
- Evans, D. J., Searles, D. J., and Rondoni, L. (2005). Application of the gallavotti-cohen fluctuation relation to thermostated steady states near equilibrium. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 71.
- Gallavotti, G. and Cohen, E. G. (1995). Dynamical ensembles in stationary states. *Journal of Statistical Physics*, 80:931–970.
- Ghil, M. and Lucarini, V. (2019). The physics of climate variability and climate change. *Reviews of Modern Physics*, 92.
- Kubo, R., Toda, M., and Hashitsume, N. (1991). Statistical mechanics of linear response. pages 146–202.
- Lorenz, E. (1963). Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20:130–141.
- Lorenz, E. (1967). The nature and theory of the general circulation of the atmosphere. *undefined*.
- Lucarini, V., Ragone, F., and Lunkeit, F. (2017). Predicting climate change using response theory: Global averages and spatial patterns. *Journal of Statistical Physics*, 166:1036–1064.

- May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467.
- Rondoni, L. and Dematteis, G. (2016). Physical ergodicity and exact response relations for low-dimensional maps. *Computational Methods in Science and Technology*, 22:71–85.
- Ruelle, D. (1998). General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium. *Physics Letters, Section A: General, Atomic and Solid State Physics*, 245:220–224.
- Saltzman, B. and Maasch, K. A. (1988). Carbon cycle instability as a cause of the late pleistocene ice age oscillations - modeling the asymmetric response. *Global Biogeochemical Cycles; (USA)*, 2:177–185.
- Searles, D. J. and Evans, D. J. (1999). Fluctuation theorem for stochastic systems. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 60:159–164.
- Searles, D. J., Rondoni, L., and Evans, D. J. (2007). The steady state fluctuation relation for the dissipation function. *Journal of Statistical Physics*, 128:1337–1363.
- Strogatz, S. H. (2018). Nonlinear dynamics and chaos : With applications to physics, biology, chemistry, and engineering. *Nonlinear Dynamics and Chaos*.