Modeling the link between carbon emissions and ocean acidification using a Lotka-Volterra dynamical system

L. F. Mendoza-Mendoza[a], I. Álvarez-Rios[b], F. Z. C. León-Altamirano[a], and F. S. Guzmán[b]

[a]*Facultad de Ciencias Físico Matemáticas de la Universidad Michoacana de San Nicolás de Hidalgo, Edificio B, Cd. Universitaria, 58040 Morelia Michoacán, México.*

[b]*Instituto de Física y Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, Edificio C-3, Cd. Universitaria, 58040 Morelia, Michoacán, México.*

Received 24 October 2024; accepted 11 March 2025

We study the possibility that the dynamics of carbon emissions and aragonite saturation in sea water can be modeled with a non-linear coupled system of two Ordinary Differential Equations. In total, there are seven parameters plus two initial conditions that we fit in order to adjust experimental data. For the fitting, we use two independent methods: Genetic Algorithms and Monte Carlo Markov Chains, both useful at fitting in high-dimensional parameter spaces. The data for the carbon emissions we deal with is obtained from several sources dedicated to monitoring the changes in these emissions over time. We calculated aragonite saturation using a combination of carbon chemistry measures from two stations. Our findings show that with this combination of ODEs and the fitting methods chosen, these two phenomena can be reproduced within an 8% error.

Keywords: Carbon emissions; ocean acidification; mathematical modeling.

DOI: https://doi.org/10.31349/RevMexFis.71.051401

1. Introduction

Since the beginning of industrialization, burning fossil fuels were used to power emerging factories, the concentration of atmospheric CO_2 has been steadily increasing. At the onset of this period, the concentration of CO_2 was approximately 280 ppm. However, current measurements indicate that the concentration has risen to ~ 425 ppm, a considerable and important level when compared to historical geological records. Analyzing ice cores from Antarctica provides an understanding of changes in carbon concentrations in the atmosphere. For the last 800,000 years, the concentration of CO_2 has never surpassed 300 ppm [1]. To find CO_2 concentration levels similar to current levels, we must go back 3 million years to the Pliocene epoch. During this period, CO_2 concentrations ranged from 365 to 415 ppm [2].

Although the concentration of CO_2 in the Pliocene epoch and the present day is similar, temperatures during the Pliocene epoch were about 3 to 4°C warmer than preindustrial levels. Additionally, sea levels were significantly higher, ranging from 5-25 meters [3]. The combination of higher sea levels and temperatures, along with elevated CO_2 concentrations, occurred during a period in Earth's history when the planet was in a state of equilibrium. This is in contrast to the current situation, where sea levels are still rising. The Pliocene epoch serves as an example of this phenomenon, although the mechanisms that led to those conditions are still being studied.

In the context of CO_2, the oceans serve as an ally by absorbing approximately 30% of anthropogenic CO_2 emissions [4]. However, this process also leads to the acidification of the oceans. To have a better explanation of this phenomenon, it is necessary to use the basis of the carbon chemistry in seawater. This chemistry can be represented by the following chemical reactions:

$$CO_{2(aq)} + H_2O \rightleftharpoons H_2CO_3, \tag{1}$$

$$H_2CO_3 \rightleftharpoons HCO_3^- + H^+, \tag{2}$$

$$HCO_3^- \rightleftharpoons CO_2^- + 2H^+, \tag{3}$$

when carbon dioxide in the atmosphere dissolves in water, it forms carbonic acid (H_2CO_3), which quickly dissociates by losing a hydrogen ion (H^+) to form two ions: bicarbonate (HCO_3^-) and carbonate (CO_2^-). These reactions are reversible and tend to reach equilibrium [5]. Under normal conditions, the exchange of CO_2 between air and sea reaches equilibrium within a year [6]. Adding more CO_2 to the atmosphere and consequently to the ocean alters the equilibrium and creates more hydrogen ions, this increase of hydrogen ions results in the ocean becoming more acidic. This excess of hydrogen ions disrupts the equilibrium in Eq. (3), and is led towards the left side, which creates more (H^+) by consuming more carbonate ions (CO_2^-).

Since the industrial era, the pH of the ocean surface water has decreased by nearly 0.1 [7]. The saturation state (Ω) determines whether the structures of carbonate minerals form or dissolve, and can be represented by the following equation [8]:

$$\Omega = \frac{[Ca^{2+}]_{\text{seawater}}[CO_3^{2-}]_{\text{seawater}}}{[Ca^{2+}]_{\text{saturation}}[CO_3^{2-}]_{\text{saturation}}}, \tag{4}$$

where *seawater* represents in situ concentrations and *saturation* refers to mineral saturation concentrations. The concentration of calcium is relatively uniform and is not changing in

the ocean, which means that the saturated state is determined by the carbonate (CO_2^-), which leaves us with the following equation:

$$\Omega_{ar} = \frac{[CO_3^{2-}]_{\text{seawater}}}{[CO_3^{2-}]_{\text{saturation}}}. \quad (5)$$

For supersaturated conditions ($\Omega > 1$) $CaCO_3$ will tend to grow, whereas for undersaturated conditions ($\Omega < 1$) will decrease. The continuous decrease of $[CO_3^{2-}]_{(\text{seawater})}$ provokes that the saturation state turns undersaturated. The saturation of aragonite (Ω_{ar}), which is a mineral form of $CaCO_3$, is a crucial component to the formation of shells and skeletons in marine organisms. Coral reefs are highly vulnerable to changes in aragonite saturation levels. A decrease in saturation can lead to coral bleaching, which is a significant issue, because coral reefs are home to approximately one-third of the world's marine fish [9].

In this paper we model the dynamics of these two variables using first order differential equations coupled through a term of interaction: 1) Carbon emissions derived from fossil based fuels and land use changes, and 2) Ocean acidification.

The reminder of the paper is organized as follows. In Sec. 2 we describe the data used and its pre-processing before fitting, as well as the mathematical model. In Sec. 3 we describe the two methods used to fit the parameters of the system, in Sec. 4 we present the fitting results, in Sec. 5 we discuss such results and in Sec. 6 we compare the context of our work with other studies and methodologies. Finally in Sec. 7 we draw some conclusions.

2. Data and model

Nowadays the leading cause of increasing carbon dioxide emissions released into the atmosphere is the use of fossil fuels. Before the industrial revolution, land-use changes were the primary cause of CO_2 emissions [10]. Although this contribution has decreased, it still remains a significant factor. This contribution, estimated to be between 22 and 43 ppm up to the year 2000 [11], is small compared to the use of fossil fuels but not insignificant. Also by the year 2000, the saturation state of aragonite has decreased to $\sim 84\%$ of pre-industrial values [12], and this percentage is expected to continue to decrease. This trend is represented in Fig. 1.

2.1. Data of CO_2 emissions

The model incorporates CO_2 emission data in the form of time series from two primary sources: fossil fuels and land-use emissions, with each source using three distinct datasets. Regarding the use of fossil fuels, the model integrates time series data from the following sources: Global Carbon Project (GCP), whose methodology and data are outlined in Ref. [13]; Community Emissions Data System (CEDS), detailed in Ref. [14] with accompanying data in Ref. [15]; and Emissions Database for Global Atmospheric Research (EDGAR), described in Ref. [16], with data sourced from [17]. These datasets are shown in Fig. 2. There are multiple estimates of CO_2 emissions due to various factors, including the emissions considered, the methodology used, and the uncertainties associated with data collection and processing. A comprehensive study in Ref. [18] discusses these factors in detail.

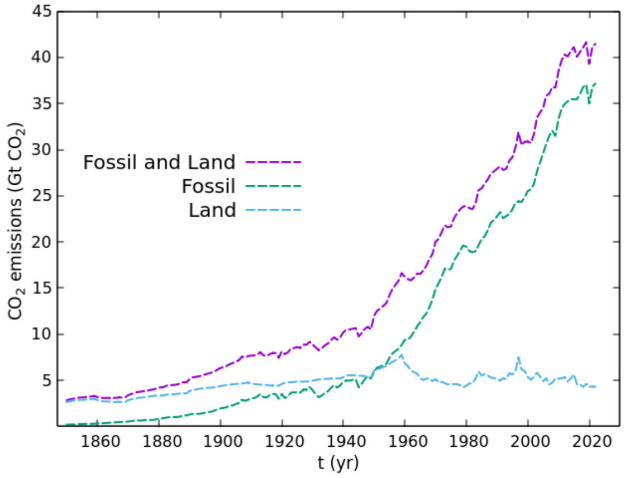

FIGURE 1. CO_2 emissions since the year 1850 from Global Carbon Project data due to land-use change, fossil fuels and the sum of both.

The flux of CO_2 emissions associated with land use change includes, but is not limited to, fluxes from deforestation, afforestation, logging and forest degradation. Two approaches for estimating the flux of CO_2 from land use change are the following: OSCAR [19] and the work of Houghton and Castanho [20], the third dataset is that of the GCP. The data of the first two approaches are reported in the GCP, a representation of these three datasets is shown in Fig. 2.

In summary, we have three sets of data of emissions due to fossil fuels and three due to land-use change. We fit the parameters of the model to the nine combinations of the data sets in Fig. 2.

2.2. Aragonite saturation data

In our analysis we use the inorganic carbon chemistry to calculate aragonite saturation data from surface seawater samples. The samples were collected from two programs: the Bermuda Institute of Ocean Sciences (BIOS) and the Hawaii Ocean Time-series (HOT) programs. BIOS encompasses two stations: Hydrostation 'S' and the Bermuda Atlantic Time-series Study (BATS). Surface data used to fit our model were obtained from [21], providing parameters such as total alkalinity (ALK), dissolved inorganic carbon (DIC), dissolved oxygen (DO), temperature, and salinity. Additionally, long-term Oligotrophic Habitat Assessment (ALOHA) data from the HOT station [22] supplied DIC, temperature, salinity, as well as the reduced isotope ratio C^{13}/C^{12} of DIC ($\delta^{13}C - DIC$) and ALK.

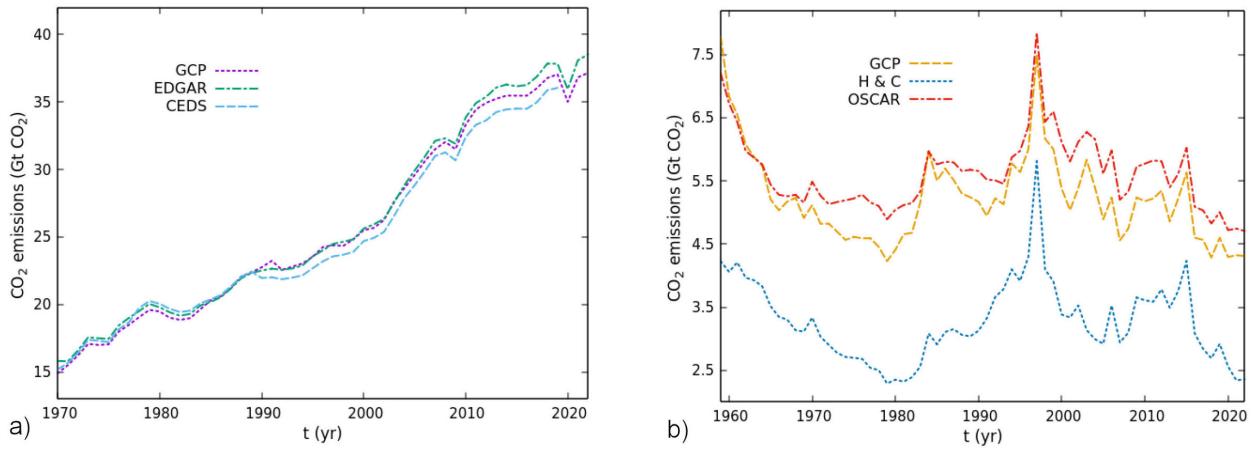

FIGURE 2. a) CO_2 emissions due to fossil fuels from three sources EDGAR, GCP and CEDS datasets. b) CO_2 emissions from the three data bases OSCAR, GCP and Houghton & Castanho, due to land-use change.

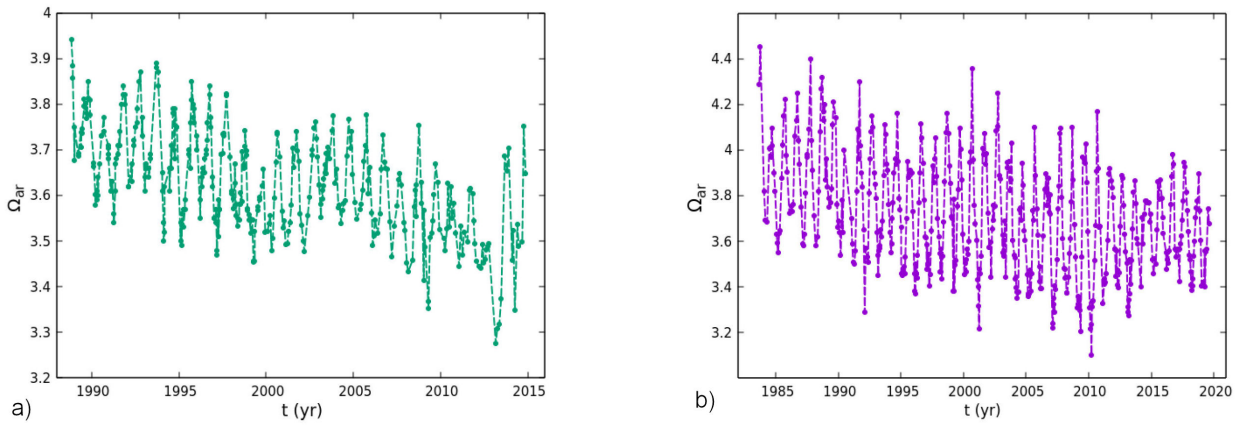

FIGURE 3. The surface aragonite saturation (Ω_{ar}) is presented as a function of time. a) The saturation calculated for the HOT station data, which has been monitoring seawater conditions since 1988. b) The resulting saturation for data from the BIOS station seawater chemistry data, which has been monitored since 1983. In both instances, aragonite saturation data was calculated using the CO2SYS software tool.

A limitation of existing data on seawater chemistry is the absence of a standardized, global time series. In contrast, there is a wealth of data on carbon obtained from various stations dedicated to this purpose, and the locations for the ocean time-series are located in coastal regions, which complicates the measurements. This challenge was further enhanced by the disruptions caused by the pandemic and the nature of the ocean. Both the HOT and BATS programs are regarded as the baseline for ocean measurements. A thorough examination of these challenges and a more effective approach to establishing a consistent ocean time series can be found in Ref. [23].

The carbon chemistry data was analyzed and aragonite saturation was determined using the CO2SYS tool. This software was originally developed for MS-DOS [24] and has since been adapted to other platforms, including MATLAB [25] and Python [26]. The version used in this study is [27]. This tool is designed to calculate parameters related to the marine carbonate system. The user inputs essential factors such as temperature, salinity, alkalinity, DIC and pH. The software uses equilibrium chemical processes and constants to calculate the quantities of DIC species, including CO_2, HCO_3^-, and CO_3^{2-}. CO2SYS provides valuable information on seawater pH, alkalinity, and carbonate chemistry, including saturation states of calcium carbonate minerals, by considering the interactions among these species. The aragonite saturation for the BATS and ALOHA datasets set are shown in Fig. 3.

In our fitting process we use the nine combinations of CO_2 emissions, in turn combined with these two data sets for aragonite saturation and finally have a total of eighteen data set combinations.

2.3. Data filtering

The data presented above can be quite intricate, as they may contain high-frequency modes that complicates modeling. However, if the amplitudes of these high-frequency modes is relatively small compared to those of the low-frequency ones, it becomes useful to filter and focus only on high-amplitude

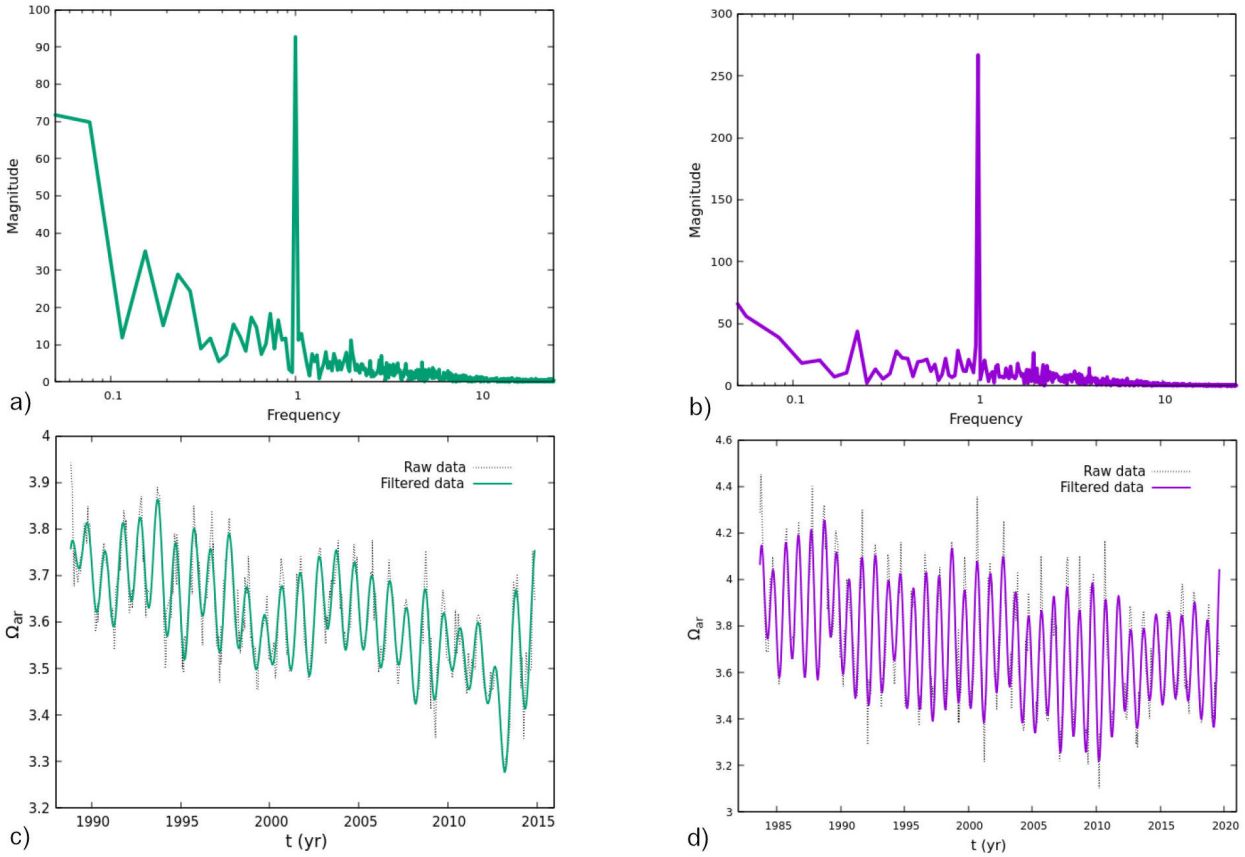

FIGURE 4. At the top we show the Fourier Transform of data sets of aragonite saturation from Fig. 3. The main peak corresponds to the dominant mode $2\pi\omega > 1$. We filter the modes with higher frequencies beyond this threshold. This filtering process eliminates subdominant high-frequency components, and allows one to focus on the dominant modes of the signal. At the bottom we show the resulting filtered data along with the original raw data from Fig. 3.

dominant modes. To achieve this, we decompose in Fourier modes calculating the discrete Fourier Transform to data, filter high frequencies and retrieve the signal with the inverse Fourier Transform.

On the top of Fig. 4 we show the Fourier Transform of data in Fig. 3 in the frequency domain. Notice that the amplitudes for $2\pi\omega > 1\text{yr}^{-1}$ are not as significant as those for lower frequencies. We then use a low-pass filter that cancels the Fourier modes of frequencies higher than the threshold $2\pi\omega_{threshold} = 1$. The filtered data are shown in the bottom part of Fig. 4.

2.4. The mathematical model

We propose a simple model based on a non-linear coupled system of equations for the two variables and investigate whether it serves to fit the data. The set of equations we employ is a system similar to the Lotka-Volterra model as follows:

$$\frac{d(\text{CO}_2)}{dt} = a(\text{CO}_2) + b(\text{CO}_2)^2 + c(\Omega_{\text{ar}})(\text{CO}_2), \qquad (6)$$

$$\frac{d(\Omega_{ar})}{dt} = -d(\Omega_{ar}) + e(\Omega_{ar})^2 - f(\Omega_{ar})(\text{CO}_2) + g\sin(\omega_A t + \phi), \qquad (7)$$

where CO_2 is the total carbon emissions and Ω_{ar} is the aragonite saturation. The linear terms in each equation

are, $a(\text{CO}_2)$, the injection rate of carbon dioxide to the oceans, and similarly $d(\Omega_{ar})$, the rate of aragonite saturation. The non-linear terms, $b(\text{CO}_2)^2$ and $e(\Omega_{ar})^2$, are non-linear terms of logistic type, that prevent exponential growth or sudden decline. The non-linear terms, $(\Omega_{ar})(\text{CO}_2)$, on the right hand side of each equation model the coupling between these two variables, g and ω_A are coefficients that introduce an oscillatory component in the aragonite equation, g determines the amplitude of the oscillation of angular frequency ω_A. The term $\sin(\omega_A t + \phi)$ is a source with a phase lag ϕ.

System (6)-(7) contains the two non-linear crossed terms with coefficients c and f, however it is expected that aragonite does not affect the emissions of CO_2 and therefore c is disregarded. In addition, it is necessary for the terms $d(\Omega_{ar})$ and $f(\Omega_{ar})(\text{CO}_2)$ to be negative due to the decreasing saturation of aragonite. The term $a(\text{CO}_2)$ must be positive as

CO_2 emissions continue to increase. The model used is finally as follows:

$$\frac{d(\mathrm{CO_2})}{dt} = a(\mathrm{CO_2}) + b(\mathrm{CO_2})^2, \tag{8}$$

$$\frac{d(\Omega_{ar})}{dt} = -d(\Omega_{ar}) + e(\Omega_{ar})^2 - f(\Omega_{ar})(\mathrm{CO_2}) + g\sin(\omega_A t + \phi). \tag{9}$$

This is the model we expect to serve as framework for understanding the correlation between climate change and ocean acidification. Then what follows is the description of the fitting method of the seven parameters in the equations and the two initial conditions that work also as parameters to be fitted.

3. Fitting methods

3.1. Genetic algorithm

The fitting consists in searching for the values of the constants a, b, d, e, f, g, ϕ and the initial conditions for the dynamical variables $\mathrm{CO_2}(0)$ and $\Omega_{ar}(0)$ in the equations of the model (8)-(9), such that the solutions have the smallest possible error.

The fact that this is a set of nine fitting parameters offers certain difficulty, added to the complexity that the equations are not fitting functions to be evaluated for each combination of trial parameters, but a system of coupled ODEs to be solved for each combination of parameters. The first strategy we implement to fit these parameters is a Genetic Algorithm (GA), whose specifications are as follows.

1. First, each individual within the population is characterized by its $DNA = (z_1, z_2, \ldots, z_{N_{\text{genes}}})$, where N_{genes} represents the total number of genes. This DNA comprises all the parameters in the model equations along with the initial conditions of the model equation $(a, b, d, e, f, g, \phi, \Omega_{ar}(0), \mathrm{CO_2}(0))$.

2. We evaluate an individual's suitability for solving the presented problem by employing a *fitness function*. Each individual, with its DNA will define a unique set of Eqs. (8)-(9). This set of equations is then solved and the resulting solution is compared with the observed data from Secs. 2.1 and 2.2. Assuming experimental data are the set $\vec{X} = (X_1, X_2, \ldots, X_N)$, and that the solution of the equations of a given individual evaluated at the same time as data are $\vec{x} = (x_1, x_2, \ldots, x_N)$, we define the fitness function as the Mean Absolute Percentage Error (MAPE):

$$f(\vec{x}, \vec{X}) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{X_i - x_i}{X_i}\right|^{-1}. \tag{10}$$

3. An initial population is constructed as the initial generation. Once generated, for simplicity we apply a *differential mutation* to the entire population since the first generation on. This mutation involves the random selection of three individuals from the population and generating a new individual using the following formula. Assuming DNA_1, DNA_2, and DNA_3 represent the DNA of three individuals with solutions $\vec{x}_1$, $\vec{x}_2$, and $\vec{x}_3$ respectively, the new individual is defined as $z_{\text{new},j} = z_{1,j} + \alpha\sum_{i=1}^{N_{\text{genes}}}(z_{2,i} - z_{3,i})$, where α is a scaling factor with a value of $\alpha = 0.5$. This new individual provides a solution $\vec{x}_{\text{new}}$. If $f(\vec{x}_{\text{new}}, \vec{X}) < f(\vec{x}_1, \vec{X})$, we replace the individual DNA_1 with DNA_{new}; otherwise, the population remains unchanged.

4. Once the original individuals have undergone mutation, the next step is to select the succeeding generation. Given a population of size $N_{\text{population}}$, we choose $N_{\text{parents}} < N_{\text{population}}$ individuals, referred to as parents, to constitute the next generation of individuals. The selection process employs the *tournament selection* method, which involves picking a random group of individuals who participate in a tournament of size $N_{\text{tournament}}$. Once the $N_{\text{tournament}}$ individuals with the highest adaptability have been determined (*i.e.*, the individuals with the greatest fitness), they are chosen to be part of the next generation. This process is repeated until all N_{parents} parents have been selected.

5. Once the parents have been selected, the subsequent step *crossover* and reproduce these individuals to obtain the remaining $N_{\text{children}} = N_{\text{population}} - N_{\text{parents}}$ individuals, referred to as children. The method for generating these children is straightforward: we randomly select two individuals from the population without repetition. From these two parents, we copy the DNA to create a child using the following procedure: we generate N_{genes} random numbers between 0 and 1. If the number in position j is less than 0.5, then element d_j is copied from the first parent; otherwise, it is copied from the second parent for $j = 1, 2, \ldots, N_{\text{genes}}$. This process is repeated until all N_{children} children are generated.

6. Now, the children have the opportunity to acquire traits that their parents do not possess through the process of *mutation*. Mutation is carried out in a manner similar to reproduction. Specifically, N_{genes} random numbers are generated. If the number at position j is less than 0.5, then the element d_j is modified by replacing it with βd_j, where β is a random number within the range of -1.1 to 1.1 for $j = 1, 2, \ldots, N_{\text{genes}}$.

7. Finally, we evolve the population during $N_{\text{generations}}$ by repeating steps 1 to 6.

3.2. MCMC method

A second strategy we use for estimating the parameters is the use of Monte Carlo Markov Chain methods. For this purpose, the Python library *emcee*, which implements an affine invariant ensemble sampling method is employed [28,29]. The specifications of the MCMC are the following:

1. *Initialization of walkers*: The walkers are initialized at random points around an initial guess. This process involves defining the number of walkers to use, ensuring a broad exploration of the parameter space. Walkers are split into two equal sets.

2. *Walker updates*: To update the position of a walker $\theta^{(n)}$ in one subset, a walker $\theta^{(j)}$ is randomly selected from the complementary set, proposing a new position using the scheme:

$$Y = \theta^{(j)} + Z \cdot (\theta^{(n)} - \theta^{(j)}), \tag{11}$$

where Z is a random variable from a distribution function $g(z)$ satisfying:

$$g(z^{-1}) = z \cdot g(z). \tag{12}$$

This distribution dictates the step size and direction.

3. *Acceptance probability*: The probability of accepting the new position is computed as:

$$\alpha = \min\left(1, Z^{D\theta-1}\frac{\pi(Y)}{\pi(\theta^{(n)})}\right), \tag{13}$$

where $D\theta$ denotes the number of parameters to estimate, and π is the log-posterior probability, computed as:

$$\pi = \ell + p_0, \tag{14}$$

with ℓ representing the likelihood, which is:

$$\ell = -\frac{1}{2}\sum\left(\frac{X_i - x_i}{x_{err}}\right)^2. \tag{15}$$

Here, x_{err} accounts for measurement errors in the data, and is defined arbitrarily as 5% of the average values of the measured variables; this term adjusts the weight of each data point according to its measurement error, ensuring that the model parameters are adjusted appropriately in response to the reliability of each measurement.

4. *Prior density function*: The prior density, $p_0(\theta)$, critical for constraining parameter values within a plausible range, is defined as:

$$p_0(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta, \\ -\infty & \text{otherwise,} \end{cases} \tag{16}$$

where Θ represents the expected range of parameter values. This function assigns a zero prior probability to parameter values outside Θ, effectively excluding them from consideration in the model fitting process.

5. *Random number generation and walker movement*: A random number U is generated from a uniform distribution. If $U \leq \alpha$ the walker moves to the proposed position Y.

6. Steps 2 to 5 are repeated for a defined number of iterations, discarding an initial percentage as the burn-in phase to ensure convergence to the stationary distribution.

4. Results

Determination of model parameters. To start with, notice that the value obtained for the natural frequency of aragonite saturation oscillation is related to the period, denoted as $T_A = 2\pi/\omega_A = 1$ yr. Essentially, this implies that the oscillation frequency of aragonite precisely covers one year, a parameter with a direct correlation to ocean temperatures. Consequently, we initially set $\omega_A = 2\pi$ yr^{-1}, the remaining coefficients and initial conditions are determined using the methods described above.

In the GA method, we employ a population of $N_{\text{population}} = 200$ individuals, with a subgroup of $N_{\text{parents}} = 50$ individuals designated as parents. We set parental selection using the tournament mechanism with size $N_{\text{tournament}} = 10$. The iterative evolution of the population was spanned a total of 2000 generations.

For the estimation with the MCMC method, a variable number of walkers, ranging from 32 to 50, were allowed to explore the parameter space across 2400 iterations. The following parameter constraints were applied to ensure that the sampled values remain within plausible ranges. Reflecting the normalized nature of these parameters, the range for parameters a, d and f were set between 0 and 1, while parameters b, e, and g were allowed to vary between -1 and 1. Constraints for ϕ and the initial conditions were set reasonably according to visual references and graphical analyses of the data. $\mathrm{CO}_2(0)$ was set to range from 18 to 32, while $\Omega_{ar}(0)$ was constrained between 3.4 and 4.8; ϕ was set between -1980 and -1992. The constraints for the initial conditions and ϕ varied according to the dataset, *e.g.*, in some cases it was more appropriate to use a range of 18 to 26 for $\mathrm{CO}_2(0)$, while in others a range of 24 to 32 was used, since applying the same restrictions and initial points for all datasets would not be meaningful.

TABLE I. List of the eighteen combinations of the six datasets of carbon emissions with the two calculated datasets of aragonite saturation. These scenarios represent the data sets fitted with the GA and MCMC methods.

Scenario	Fossil	Land-use	Aragonite
1	EDGAR	OSCAR	BATS
2	EDGAR	H & C	BATS
3	EDGAR	GCP	BATS
4	GCP	OSCAR	BATS
5	GCP	H & C	BATS
6	GCP	GCP	BATS
7	CEDS	OSCAR	BATS
8	CEDS	H & C	BATS
9	CEDS	GCP	BATS
10	EDGAR	OSCAR	ALOHA
11	EDGAR	H & C	ALOHA
12	EDGAR	GCP	ALOHA
13	GCP	OSCAR	ALOHA
14	GCP	H & C	ALOHA
15	GCP	GCP	ALOHA
16	CEDS	OSCAR	ALOHA
17	CEDS	H & C	ALOHA
18	CEDS	GCP	ALOHA

TABLE II. Parameters and initial conditions for the eighteen scenarios of CO_2 emissions and aragonite saturation determined with the GA method.

	Parameters								
Scenario	a $\times 10^{-2}$	b $\times 10^{-5}$	d $\times 10^{-4}$	e $\times 10^{-4}$	f $\times 10^{-11}$	g	ϕ	$CO_2(0)$	$\Omega_{ar}(0)$
1	1.78062	-6.71319	14.1911	-2.37143	89.08946	-0.38604	-1983.7758	25.39637	4.1386
2	1.92565	-5.37623	0.51377	-6.12854	35.70578	-0.38707	-1983.77506	22.71628	4.14274
3	1.06685	14.67125	0.132	-6.19577	2.62082	-0.38734	-1983.77521	25.27551	4.14179
4	1.641	-4.17001	21.9954	-0.33098	0.12335	-0.38638	-1983.77461	25.52286	4.13918
5	1.7381	-1.1915	0.67985	-6.07091	0.14086	-0.38763	-1983.77507	22.85301	4.1427
6	1.2548	6.61093	12.84953	-2.76727	6.24815	-0.38772	-1983.77526	25.39623	4.14075
7	2.11332	-19.5204	4.85694	-4.94869	80.11934	-0.38752	-1983.77616	25.87321	4.14211
8	1.5541	1.16003	16.18303	-1.92939	0.21685	-0.38779	-1983.77477	22.81704	4.14185
9	1.1141	9.20088	12.4778	-2.9088	0.26758	-0.38728	-1983.7751	25.14678	4.14123
Scenario	a $\times 10^{-2}$	b $\times 10^{-4}$	d $\times 10^{-3}$	e $\times 10^{-3}$	f $\times 10^{-4}$	g	ϕ	$CO_2(0)$	$\Omega_{ar}(0)$
10	1.3652	1.4411	3.49188	2.46133	2.3697	-0.1535	-1988.33277	26.56956	3.79583
11	0.16077	5.68478	0.48754	0.003015	0.66757	-0.15688	-1988.34453	24.81157	3.82175
12	1.37276	1.41645	6.22063	2.06561	1.13056	-0.15677	-1988.33436	26.18308	3.8062
13	1.08842	1.98012	2.3986	0.85305	0.9621	-0.15714	-1988.34043	26.7974	3.8125
14	0.04962	5.9923	2.2185	0.99352	1.26365	-0.1546	-1988.3489	25.06447	3.8112
15	1.61054	0.38942	6.59092	2.22466	1.22588	-0.15627	-1988.35184	26.30771	3.81151
16	1.6427	0.43744	7.17167	5.18125	4.34882	-0.14928	-1988.30317	25.87427	3.76532
17	1.61833	0.86618	29.61073	7.46365	0.00000771	-0.15613	-1988.34528	23.89070	3.81352
18	1.62703	0.43333	0.0000408	-0.56901	0.15869	-0.1567	-1988.31606	25.51681	3.824

TABLE III. The parameters and initial conditions for the eighteen scenarios of CO_2 emissions and aragonite saturation, determined with the MCMC method.

	Parameters								
Scenario	a $\times 10^{-2}$	b $\times 10^{-4}$	d $\times 10^{-3}$	e $\times 10^{-4}$	f $\times 10^{-6}$	g	ϕ	$CO_2(0)$	$\Omega_{ar}(0)$
1	1.99887	-1.14315	0.77254	-4.565	2.41097	-0.9803	-1983.56757	24.96869	4.06022
2	2.16043	-1.21597	3.04979	1.58082	3.79969	-0.99025	-1983.63779	22.5052	4.0653
3	1.60143	-0.001003	11.6424	24.41474	3.68914	-0.99929	-1983.63153	24.83085	4.0547
4	2.0119	-1.44896	1.10916	-2.9652	4.68091	-0.9983	-1983.56689	25.18491	4.05418
5	2.26516	-1.84192	7.51611	13.95453	2.2774	-0.99733	-1983.71097	22.65008	4.03752
6	1.68305	-0.49783	5.23518	9.49898	21.87696	-0.99431	-1983.56459	24.9861	4.05041
7	1.41426	0.187367	7.49769	15.47677	4.94975	-0.99544	-1983.64748	25.08936	4.00907
8	1.58831	0.15385	4.18502	6.61714	15.52485	-0.9848	-1983.63816	22.60821	4.03491
9	0.99975	1.44601	11.09472	24.15899	3.2664	-0.99539	-1983.67777	24.91599	4.03878
Scenario	a $\times 10^{-4}$	b $\times 10^{-4}$	d $\times 10^{-2}$	e $\times 10^{-3}$	f $\times 10^{-5}$	g	ϕ	$CO_2(0)$	$\Omega_{ar}(0)$
10	14.28837	4.89799	2.35865	6.41611	5.78787	0.48753	-1986.9857	27.07203	3.77466
11	0.06799	6.18215	0.66379	2.07707	11.67232	-0.59398	-1990.33776	24.82025	3.8124
12	0.10755	5.42964	1.97458	5.3414	8.0648	-0.58278	-1984.0429	26.66822	3.82702
13	0.51715	5.0431	4.57584	12.13521	2.70195	-0.4858	-1990.30553	27.27379	3.78831
14	0.08591	5.88433	7.89313	21.06526	0.33066	-0.60438	-1990.1263	24.98474	3.78982
15	0.87211	5.12049	0.58529	1.77675	9.00662	-0.56443	-1990.07467	26.84374	3.79556
16	17.33548	4.68801	1.8036	5.44236	13.27927	0.5708	-1986.98291	26.51143	3.79849
17	0.33356	6.08583	4.22089	1.19433	10.68011	-0.54074	-1990.02713	24.27306	3.76326
18	1.15345	5.29483	1.76315	5.37265	14.43337	-0.60101	-1984.04978	26.11156	3.81525

Data set to fit. The combination of the six datasets related to CO_2 emissions, three of which are attributed to fossil fuels and the remaining three to land-use change, results in a total of nine combinations. Furthermore, the incorporation of the two datasets for aragonite saturation expands the combination to eighteen scenarios for the Lotka-Volterra model, as explicitly described in Table I.

Finally, the results obtained from the fittings with the GA are presented in Table II, whereas those using the MCMC method are contained in Table III. We discuss the results in the following section.

5. Discussion of the results

The first expectation from a problem with nine parameters is that the chances to be a degenerate problem are high. This means that various combinations of the parameters can imply fittings of data with similar error but different combinations of the fitting parameters. The plots of how the fittings look like for a few cases, specifically the combinations 3, 7, 12, 16 from the Tables, are shown in Fig. 5. The errors achieved by the two methods are of the same order and the fitting curves are very similar. Notice from Tables II and III, that fitting parameters differ, even though the two methods have similar errors.

The evolution of error of the fitting with the GA appears in Fig. 6 for the 18 combinations of data sources. In all cases the error approaches a minimum with values lower than 8%.

On the other hand, the convergence of the MCMC method toward attractor values of the fitting parameters is shown in Fig. 7, where the walker trajectories for Scenario 2 are portrayed.

Concerning the values of the parameters, what calls the attention more is the discrepancy of the constant f, that couples CO_2 and Ω_{ar} in the evolution Eq. (9). Using the GA it happens that using data from BATS and those from ALOHA, this parameter changes by seven orders of magnitude. This indicates that data from ALOHA allow a better coupling, at least using the GA method found this local minimum in the parameter space. This discrepancy is of one order of magnitude when fitting the data with the MCMC method. Other than that, the values of other parameters do not change as dramatically.

6. Context of our model and approach

Before describing our conclusions, we describe where our analysis can be located within the state of the art in the subject. As mentioned before, the lack of a uniform and consistent global time series of ocean trends leads to most studies focusing on specific regions.

Another aspect to consider is the methodologies employed, since the use of ordinary differential equations (ODEs) is not the norm.

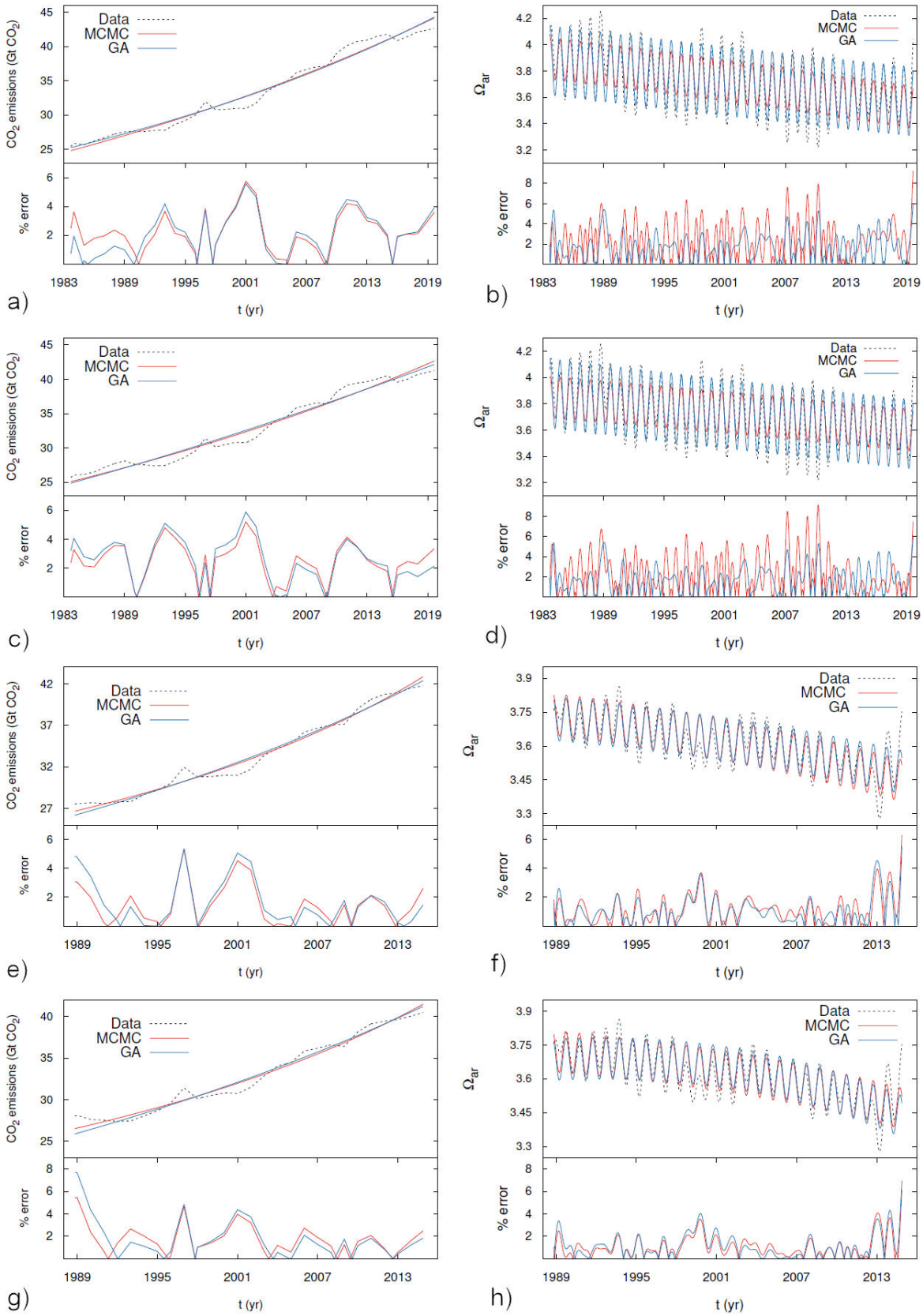

FIGURE 5. As a sample of the fittings, from top to bottom we show data and fittings of scenarios 3, 7, 12 and 16. Dotted black lines correspond to data, blue/red to the fitting models with the GA/MCMC methods.

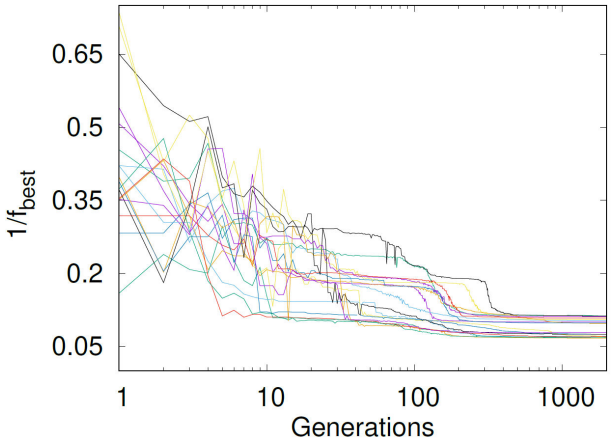

FIGURE 6. Error of the fitting using the GA method as function of time, as the reciprocal of fitness of the most fitted member of each generation. We show the normalized reciprocal fitness for the 18 combinations of data sources.

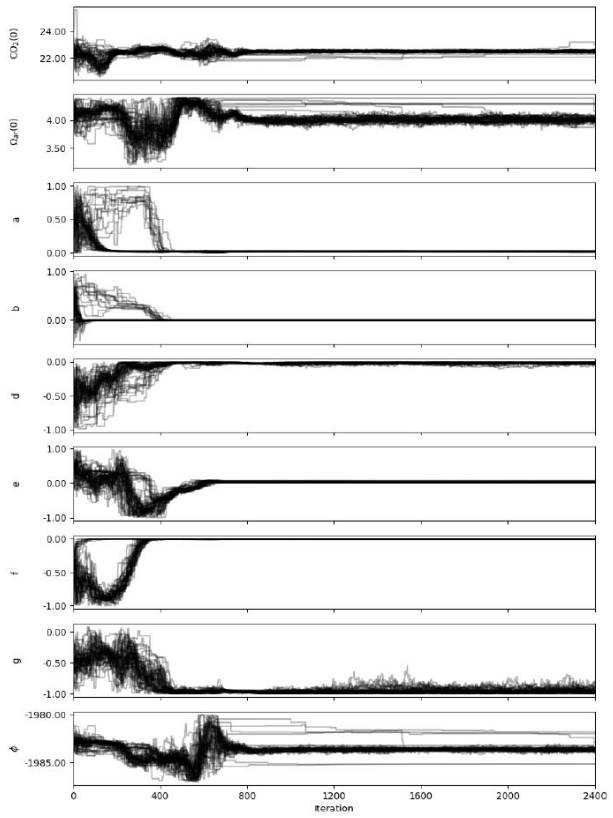

FIGURE 7. Each panel shows the value of each parameter as function of time and demonstrate how it stabilizes. The lines appear thick because they represent the trajectories of all the walkers together. This figure, corresponding to Scenario 2, also highlights the burn-in stage of the process, characterized by the spiky behavior before approximately 800 steps.

The following are some examples of the previously mentioned statements. One approach uses biogeochemical models, which prioritize NPZD in conjunction with carbon modules. This integration correlates the interaction between ocean dynamics and a more biological approach, with the incorporation of carbon modules enabling the modeling of ocean acidification. Illustrative examples of this methodology can be found in a study conducted in the Northwestern Pacific [30], and another one in the Bering Sea [31]. It is interesting that both the NPZD and the carbon modules are modeled with uncoupled differential equations.

Concerning numerical methods, a multi-linear regression is applied to the BATS data to forecast ocean acidification up to the year 2100, for a number of scenarios [32]. The use of deep learning to estimate changes related to ocean acidification in the Atlantic Ocean is presented in Ref. [33]. In these last two cases, no differential equations are used.

The framework we propose lies within the context of the above-mentioned examples, centered in the use of ordinary differential equations (ODEs) to model data in conjunction with two optimization techniques. This approach offers a distinct advantage over NPZD models, which necessitate the estimation of numerous biological parameters. Furthermore, the Lotka-Volterra system we use, can be applied to any ocean time series, irrespective of its geographical location.

7. Conclusions

We have proposed a system of coupled ODEs as a model that can simultaneously fit data for the evolution of CO_2 emissions and the aragonite saturation Ω_{ar}. In order to test the model, we used data from various sources of these quantities and fitted the combined data using two different methods. First using a Genetic Algorithm with a set of general evolution rules that approaches a minimum of the error, and second, using a Markov Chain Monte Carlo method that also finds a local minimum of the error.

We show the model is a suitable candidate for explaining the data and the future behavior of these two variables. Despite of its simplicity, it is able to find a corelation between CO_2 emissions and aragonite saturation makes, which makes it a usable framework for analyzing their behavior.

Concerning most of the parameters of the system of ODEs, apparently the two quantities can be ruled by independent equations, namely (8) and (9), as long as the coefficient f is pretty small, because it represents the only coupling between the two variables modeled. The result of this study indicates that the modeling of each variable as an independent entity is advantageous. While this finding is beneficial, it is a consequence of the parameter estimation and the degenerate nature of the problem under analysis.

The GA method found a local minimum where the coupling is important, seven orders of magnitude higher when using data from ALOHA than when using data from BATS.

We believe that the model, tested as we have done, is interesting and can be generalized to include models of other quantities that characterize the global climate on Earth.

8. Data availability

This document does not provide the data directly. Instead, the tables contain the fitting parameters that can be used to solve the equations of the model 8-9 and recreate the charts.

Regarding the data utilized in this work, we wish to inform the reader that the data employed in this study is available at the following URLs: https://edgar.jrc.ec.europa.eu/report_2023 and https://edgar.jrc.ec.europa.eu/dataset_ghg80. The CEDS data is accessible via the Zenodo repository at the following URL: https://zenodo.org/records/4741285. Additional datasets pertinent to carbon emissions can be obtained as supplementary files in Ref. [13]. The supplementary file [21] contains the observed seawater data for the BATS and Hydrostation 'S'. Information regarding the ALOHA station can be accessed at https://scrippsco2.ucsd.edu/data/seawater_carbon/ocean_time_series.html.

Acknowledgments

LFMM and IAR receive support from CONAHCyT graduate scholarship program. FZCLA receives support from CONAHCyT through the SNI3 scholarship program. This research is supported by grant CIC-UMSNH-4.9.

1. D. Lüthi *et al.*, High-resolution carbon dioxide concentration record 650,000800,000 years before present, *Nature* **453** (2008) 379, https://doi.org/10.1038/nature06949
2. M. Pagani, Z. Liu, J. Lariviere, and A. C. Ravelo, High earth-system climate sensitivity determined from pliocene carbon dioxide concentrations, *Nature Geoscience* **3** (2010) 27, https://doi.org/10.1038/ngeo724
3. Intergovernmental Panel on Climate Change (IPCC), Framing, context, and methods, in Climate Change 2021 - The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge University Press, 2023) p. 147-286.
4. D. Chen *et al*., Framing, context, and methods, in Climate Change 2021: The Physical Science Basis. Contribution of working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change ed V Masson-Delmotte *et al.* Change (Cambridge University Press, 2021) p. 147- 286.
5. F. J. Millero *et al*., Dissociation constants for carbonic acid determined from field measurements, *Deep Sea Research Part I: Oceanographic Research Papers* **49** (2002) 1705, https://doi.org/10.1016/S0967-0637(02)00093-6
6. S. C. Doney, V. J. Fabry, R. A. Feely, and J. A. Kleypas, Ocean acidification: The other CO2 problem, *Annual Review of Marine Science* **1** (2009) 169, https://doi.org/10.1146/annurev.marine.010908.163834.
7. C. Garcia-Soto *et al.*, An overview of ocean climate change indicators: Sea surface temperature, ocean heat content, ocean ph, dissolved oxygen concentration, arctic sea ice extent, thickness and volume, sea level and strength of the amoc (atlantic meridional overturning circulation), *Frontiers in Marine Science* **8** (2021), https://doi.org/10.3389/fmars.2021.642372
8. S. Barker and A. J. Ridgwell, Ocean acidification, *Nature Education Knowledge* **3** (2012) 10.
9. J. M. Lough and M. J. H. van Oppen, Introduction: Coral bleaching-patterns, processes, causes and consequences, in Coral Bleaching: Patterns, Processes, Causes and Consequences, edited by Madeleine J. H. van Oppen and Janice M. Lough (Springer International Publishing, Cham, 2018) pp. 1-8. https://doi.org/10.1007/978-3-319-75393-51
10. J. G. Canadell *et al.*, Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change (Cambridge University Press, 2021).
11. V. Brovkin, S. Sitch, W. Von Bloh, M. Claussen, E. Bauer, and W. Cramer, Role of land cover changes for atmospheric Co2 increase and climate change during the last 150 years, *Global Change Biology* **10** (2004) 1253, https://doi.org/10.1111/j.1365-2486.2004.00812.x
12. L.-Q. Jiang, R. A. Feely, B. R. Carter, D. J. Greeley, D. K. Gledhill, and K. M. Arzayus, Climatological distribution of aragonite saturation state in the global oceans, *Global Biogeochemical Cycles* **29** (2015) 1656, https://doi.org/10.1002/2015GB005198
13. P. Friedlingstein *et al*., Global carbon budget 2023, *Earth System Science Data* **15** (2023) 5301, https://doi.org/10.5194/essd-15-5301-2023.
14. R. M. Hoesly *et al.*, Historical (1750-2014) anthropogenic emissions of reactive gases and aerosols from the community emissions data system (ceds), *Geoscientific Model Development* **11** (2018) 369, https://doi.org/10.5194/gmd-11-369-2018
15. Patrick R O'Rourke *et al.*, Ceds v 2021 04 21 release emission data (_v_2021_02_05) [Dataset] (Zenodo, 2021).
16. G. Janssens-Maenhout *et al.*, Edgar v4.3.2 global atlas of the three major greenhouse gas emissions for the period 1970-2012, *Earth System Science Data* **11** (2019) 959, https://doi.org/10.5194/essd-11-959-2019.
17. European Commission, Joint Research Centre (JRC), International Energy Agency (IEA), EDGAR (Emissions Database for Global Atmospheric Research) Community GHG Database, (2023), a collaboration between the European Commission, Joint Research Centre (JRC), the International Energy Agency (IEA), and comprising IEA-EDGAR CO2, EDGAR CH4, EDGAR N2O, EDGAR F-GASES version 8.0.
18. R. M. Andrew, A comparison of estimates of global carbon dioxide emissions from fossil carbon sources, *Earth System Science Data* **12** (2020) 1437, https://doi.org/10.5194/essd-12-1437-2020

19. T. Gasser, L. Crepin, Y. Quilcaille, R. A. Houghton, P. Ciais, and M. Obersteiner, Historical CO2 emissions from land use and land cover change and their uncertainty, *Biogeosciences* **17** (2020) 4075, https://doi.org/10.5194/bg-17-4075-2020

20. R. A. Houghton and A. Castanho, Annual emissions of carbon from land use, land-use change, and forestry from 1850 to 2020, *Earth System Science Data* **15** (2023) 2025, https://doi.org/10.5194/essd-15-2025-2023

21. N. R. Bates and R. J. Johnson, Acceleration of ocean warming, salinification, deoxygenation and acidification in the surface subtropical North Atlantic Ocean, *Communications Earth and Environment* **1** (2020) 33, https://doi.org/10.1038/s43247-020-00030-5,

22. T. J. Lueker, C. D. Keeling, P. R. Guenther, M. Whalen, and W. G. Mook, Inorganic Carbon Variations in Surface Ocean Water near Bermuda, Tech. Rep. (UC San Diego: Scripps Institution of Oceanography, 1998).

23. A. J. Sutton *et al.*, Advancing best practices for assessing trends of ocean acidification time series, *Frontiers in Marine Science* **9** (2022), https://doi.org/10.3389/fmars.2022.1045667.

24. E. Lewis, D. Wallace, and L. J. Allison, Program developed for CO2 system calculations, (1998), https://doi.org/10.2172/639712.

25. J. D. Sharp *et al*., Co2sysv3 for matlab, (2020).

26. M. P. Humphreys, E. R. Lewis, J. D. Sharp, and D. Pierrot, Pyco2sys v1.8: marine carbonate system calculations in python, *Geoscientific Model Development* **15** (2022) 15, https://doi.org/10.5194/gmd-15-15-2022

27. D. Pierrot, J.-M. Epitalon, J. C. Orr, E. Lewis, and D. W. R. Wallace, Ms excel program developed for CO2 system calculations - version 3.0, GitHub repository (2021).

28. D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, emcee: The mcmc hammer, *Publications of the Astronomical Society of the Pacific* **125** (2013) 306, https://doi.org/10.1086/670067.

29. J. Goodman and J. Weare, Ensemble samplers with affine invariance, *Communications in Applied Mathematics and Computational Science* **5** (2010) 65, http://dx.doi.org/10.2140/camcos.2010.5.65

30. M. Ishizu, Y. Miyazawa, T. Tsunoda, and X. Guo, Development of a biogeochemical and carbon model related to ocean acidification indices with an operational ocean model product in the north western pacific, *Sustainability* **11** (2019) 2677, https://doi.org/10.3390/su11092677.

31. D. J. Pilcher *et al*., Modeled effect of coastal biogeochemical processes, climate variability, and ocean acidification on aragonite saturation state in the bering sea, *Frontiers in Marine Science* **5** (2019), https://doi.org/10.3389/fmars.2018.00508.

32. N. Ben Hadid, V. Guglielmi, F. Touratier, A. Mouakher, and C. Goyet, Ocean acidification from 1955 to 2100 at bats and hydrostation s (bermuda area), (2024), preprint, available at SSRN. https://dx.doi.org/10.2139/ssrn.4779316.

33. C. Galdies and R. Guerra, High resolution estimation of ocean dissolved inorganic carbon, total alkalinity and ph based on deep learning, *Water* **15** (2023), https://doi.org/10.3390/w15081454.