

Deep Learning and Dynamical Systems Approaches to Critical Transitions in Socio–Climate and Complex Systems

by

Yazdan Babazadeh Maghsoodlo

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Applied Mathematics

Waterloo, Ontario, Canada, 2025

© Yazdan Babazadeh Maghsoodlo 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Iain Moyles
Professor, Dept. of Mathematics and Statistics, York University

Supervisor(s): Chris Bauch
Professor, Dept. of Applied Mathematics, University of Waterloo

Madhur Anand
Professor, School of Environmental Sciences, University of Guelph

Internal Members: Jun Liu
Professor, Dept. of Applied Mathematics, University of Waterloo

Barbara Zemskova
Professor, Dept. of Applied Mathematics, University of Waterloo

Internal-External Member: Eihab Abdel-Rahman
Professor, Dept. of Systems Design Engineering, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Statement of Contributions

- Chapter 2: The work in this chapter was performed by Yazdan Babazadeh Maghsoodlo (YBM) under the supervision of Dr. Madhur Anand (MA) and Dr. Chris T. Bauch (CTB), and resulted in a co-authored publication in *Mathematics* 2025, 13(17), 2782. Contributions were as follows: MA and CTB conceived the study. YBM developed the methodology, implemented the software, carried out the validation, conducted the formal analysis and investigation, curated the data, and wrote the first draft of the manuscript. All authors contributed to the revision of the manuscript. All authors approved the final published version.
- Chapter 3: The work in this chapter was performed by Yazdan Babazadeh Maghsoodlo (YBM) under the supervision of Dr. Madhur Anand (MA) and Dr. Chris T. Bauch (CTB), and resulted in a co-authored preprint available on arXiv:2509.04683. Contributions were as follows: MA and CTB conceived the study. YBM developed the methodology, implemented the software, carried out the validation, conducted the formal analysis and investigation, curated the data, and wrote the first draft of the manuscript. All authors contributed to the revision of the manuscript. All authors approved the final version.
- Chapter 4: The work in this chapter was performed by Yazdan Babazadeh Maghsoodlo (YBM) under the supervision of Dr. Madhur Anand (MA) and Dr. Chris T. Bauch (CTB), and resulted in a co-authored publication in *Proceedings of the Royal Society A*, Volume 481, Issue 2322, Article 20250376. Contributions were as follows: MA and CTB conceived the study. YBM developed the methodology, implemented the software, performed the data curation, conducted the formal analysis and investigation, carried out the validation, generated the visualizations, and wrote the first draft of the manuscript. All authors contributed to the revision of the manuscript and approved the final published version.
- Chapter 5: The work in this chapter was performed by Yazdan Babazadeh Maghsoodlo (YBM) under the supervision of Dr. Madhur Anand (MA) and Dr. Chris T. Bauch (CTB), and resulted in a co-authored preprint available on arXiv:2509.25111. Contributions were as follows: MA and CTB conceived the study. YBM developed the methodology, implemented the software, performed the data curation, conducted the formal analysis and investigation, carried out the validation, generated the visualizations, and wrote the first draft of the manuscript. All authors contributed to the revision of the manuscript and approved the final version.

Abstract

This thesis explores how dynamical systems, stochastic processes, and deep learning can be integrated to study critical transitions in socio-climate and other complex systems. Chapter 1 establishes the conceptual foundation, introducing complex systems, tipping points, bifurcation theory, stochasticity, early warning signals, and the role of deep learning. It also highlights flickering as a precursor to collapse and motivates the importance of coupled socio-climate feedbacks. Chapter 2 develops a hybrid CNN–LSTM framework to classify bifurcations in noisy time series. Trained on synthetic dynamical models, the classifier generalises to empirical data and outperforms traditional early warning signals, offering a robust method to identify fold, Hopf, and transcritical bifurcations. Chapter 3 introduces a deep learning approach to detect flickering dynamics, noise-driven switching between alternative equilibria. The model distinguishes true flickering from noise-induced variance inflation across diverse systems and demonstrates applicability to empirical data such as palaeoclimate records and physiological signals, providing an early warning beyond variance-based methods. Chapter 4 presents a coupled socio-climate model where social behaviour feeds back on emissions and climate thresholds. Results show that social dynamics, such as faster learning rates or stronger norms, can delay or prevent climate tipping, while delays or weak norms accelerate collapse. This chapter highlights the potential of social tipping points to stabilize climate trajectories. Chapter 5 evaluates whether binary opinion models suffice to represent socio-climate interactions compared to richer spectrum models. Using replicator and Friedkin–Johnsen frameworks coupled to climate-carbon and forest-grassland systems, the study finds that binary models capture essential coupled dynamics with surprising accuracy, despite their simplicity. Together, the chapters demonstrate that combining dynamical systems theory, stochastic analysis, and deep learning yields powerful tools to anticipate tipping points. The findings advance both methodological development and practical insight, showing that human social responses can critically shape whether climate transitions are mitigated or exacerbated.

Acknowledgements

I am sincerely thankful to my supervisors, Professor Chris Bauch and Professor Madhur Anand, for their constant support and guidance throughout the years of my PhD program. This thesis would not have been possible without them.

Dedication

*To my wife Tina,
my parents, Jalal and Farzaneh,
my brother Alborz,
and my parents-in-law, Reza and Susan.*

Table of Contents

Examining Committee	ii
Author's Declaration	iii
Statement of Contributions	iv
Abstract	v
Acknowledgements	vi
Dedication	vii
List of Figures	xii
List of Tables	xix
1 Introduction	1
1.1 Complex Systems and Tipping Points	1
1.2 Dynamical Systems and Bifurcation Theory	3
1.3 Stochasticity and Noise	8
1.4 Early Warning Signals (EWS)	10
1.5 Deep Learning for Dynamical Systems	11
1.6 Flickering as a Precursor of Collapse	16
1.7 Why These Topics Matter	22
1.7.1 Aims and Objectives	23
1.7.2 Thesis Outline	24
2 Deep Learning for Bifurcation Detection	26
2.1 Abstract	27

2.2	Introduction	27
2.3	Methods	29
2.3.1	Constructing training data set	29
2.3.2	Model Selection Rationale	35
2.4	Results	38
2.5	Discussion	46
3	Echoes Before Collapse	48
3.1	Abstract	49
3.2	Introduction	49
3.3	Deep learning model and dataset construction	52
3.3.1	Details of Training Dataset	52
3.4	Results	55
3.4.1	Details of Simulated Models	57
3.4.2	Generalization Across Dynamical Systems	58
3.4.3	Quantitative Evaluation via ROC Analysis	60
3.4.4	Application to Empirical Data	61
3.5	Discussion	63
4	Social dynamics can delay or prevent climate tipping points	65
4.1	Abstract	66
4.2	Introduction	66
4.3	Model Description	68
4.3.1	Behaviour Dynamics	68
4.4	Earth system model	70
4.4.1	Atmospheric carbon dynamics	70
4.4.2	Full Earth system model equations	71
4.4.3	Biogeochemical fluxes	71
4.4.4	Radiative forcing and temperature	72
4.4.5	Simulation	73
4.5	Results	73
4.6	Discussion	81
5	When Simple is Enough: Binary Models Capture Social Complexity in Coupled Human–Environment Systems	83
5.1	Abstract	84
5.2	Introduction	84
5.3	Methods	86

5.3.1	Forest–grassland mosaic model	86
5.3.2	Climate–carbon model	87
5.3.3	Replicator social model	87
5.3.4	FJ social model	88
5.3.5	Comparison Metric	89
5.4	Results	91
5.4.1	Coupling Between the Replicator Social Model and the Climate–Carbon Model	91
5.4.2	Coupling Between the FJ Social Model and the Climate–Carbon Model	93
5.4.3	Coupling of the Forest Model and the Replicator Social Model	96
5.4.4	Coupling of the Forest Model and the FJ Social Model	98
5.5	Discussion	101
6	Conclusion	102
	References	104
	APPENDICES	125
A		126
A.1	Deep Learning Model Architecture	126
A.2	Validation on Bifurcating and Null Dynamical Systems	127
B		131
B.1	Deep Learning Architecture	131
B.2	Data Availability	132
C		133
C.1	Stability Analysis of the social model	133
C.2	Sensitivity of Climate Impact to Social Learning Rate	134
C.3	Table. Labels for state variables and dynamic processes in the socio-climate model.	136
C.4	Sensitivity Analysis	137
C.5	Historical record of CO_2 emission	139
C.6	Data and Code Availability	139
D		140
D.1	Formulation of Climate–Carbon Model	140
D.2	Full details of the replicator–climate–carbon model coupling	141

D.3 Full details of the FJ social model–forest cover coupling	142
D.4 Data Availability	146

List of Figures

1.1	Empirical anoxia dataset (64PE) showing a tipping point, with the vertical dashed line indicating its approximate location.	2
1.2	Bifurcation dynamics in three canonical models: Rosenzweig–MacArthur with a transcritical bifurcation (left), Rosenzweig–MacArthur with a Hopf bifurcation (middle), and May’s harvesting model with a fold bifurcation (right). Vertical dashed lines indicate the approximate locations of the transitions.	4
1.3	Loss of resilience near a bifurcation: as the system approaches the tipping point, recovery from perturbations slows and autocorrelation increases. Image taken from [125]	6
1.4	Change in variance and lag-1 autocorrelation in fold, transcritical, and Hopf systems, illustrating their role as early warning signals of bifurcations. Image taken from [34]	7
1.5	Illustration of a convolutional neural network (CNN). Convolutional layers extract local features through filtering operations, pooling layers reduce dimensionality, and fully connected layers integrate the learned representations to perform classification or prediction. The image is taken from [6].	13
1.6	Illustration of a long short-term memory (LSTM) network, highlighting its memory cells and gating mechanisms for capturing long-range temporal dependencies. The image is taken from [107].	15

1.7	Comparison of dynamical regimes leading to transitions. (a) Classical critical slowing down: as the system approaches a bifurcation, recovery from perturbations slows, increasing variance and autocorrelation. (b) Driver-mediated flickering: slow changes in external forcing cause repeated transitions between coexisting states. (c) Noise-mediated flickering: random perturbations alone trigger state switching even without parameter drift. Corresponding panels (a1–c3) show the evolution of variance, lag-1 autocorrelation (AR1), and skewness computed over sliding windows. Image adapted from [45].	17
1.8	Schematic diagram of feedbacks between the social and climate systems, mediated by emissions, regional impacts, and behavioral responses. Image adapted from [18].	20
2.1	(Left) Model performance as a function of training data fraction. (Right) Precision as a function of time series length.	35
2.2	DL model providing EWS prior to bifurcation points for Rosenzweig–MacArthur and May’s harvesting model. (A–C) Synthetic time series from three canonical bifurcation models with additive coloured noise: (A) Rosenzweig–MacArthur predator-prey system near a transcritical bifurcation; (B) Rosenzweig–MacArthur with Hopf bifurcation; (C) May’s harvesting model near a fold bifurcation. (D–F) DL model predictions for the systems in (A–C), showing temporal evolution of bifurcation class probabilities. Probabilities are represented with distinguishable markers and line styles for print clarity: triangle (fold), square (transcritical), circle (Hopf), and x (null). The vertical dashed line indicates the approximate location of the tipping point.	38
2.3	DL model providing EWS prior to bifurcation points for Van der Pol, SIS, and Neural activation model. (A–C) Synthetic time series from three canonical bifurcation models with additive white noise: (A) Van der Pol oscillator approaching a Hopf bifurcation; (B) SIS model undergoing a transcritical bifurcation; (C) Neural activation model with a fold bifurcation. The vertical dashed lines indicate the location of the bifurcation. (D–F) Output of the DL classifier showing predicted bifurcation probabilities over time for the corresponding systems in (A–C). Distinct line styles and markers denote bifurcation types: triangle (fold), square (transcritical), circle (Hopf), and x (null).	41

2.4	DL model providing EWS before bifurcation points. (A–D) Real-world time series examples: (A–B) Thermoacoustic instability signals from combustor experiments at ROF 5 and ROF 10; (C–D) Geological anoxia indicators from sediment cores at two locations. (E–H) Corresponding DL-based bifurcation predictions for (A–D), with line styles and markers indicating the inferred probability of each class. This illustrates how the model generalizes from synthetic to empirical data. Distinct line styles and markers denote bifurcation types: triangle (fold), square (transcritical), circle (Hopf), and x (null). The vertical dashed lines are the approximate location of the tipping point.	43
2.5	AUC scores of White Noise and Red Noise classifiers across six mathematical models Classifier performance across six dynamical systems under varying levels of noise correlation (redness). Each subplot displays the AUC score as a function of redness for white noise conditions (WNC, blue circles) and red noise conditions (RNC, orange squares). Top row: Rosenzweig–MacArthur model with transcritical and Hopf bifurcations, and May’s harvesting model with a fold bifurcation. Bottom row: SIS model (transcritical), Van der Pol oscillator (Hopf), and neural activation model (fold). The results highlight that while most systems retain high classification accuracy across redness levels, performance in certain fold bifurcations, especially under WNC in the neural activation model and under RNC in the May model, shows degradation as temporal correlation increases.	44
3.2	Deep learning model architecture for flickering pattern detection. The CNN–LSTM framework detects flickering patterns as early-warning indicators of tipping points. Simulated time series with and without flickering (parameterized by p_0 and p^*) are preprocessed (see Box 1) and used to train the model, which outputs a binary classification indicating the presence of flickering.	55
3.3	Deep-learning discrimination of true flickering versus noise-modulated null dynamics across six models. For each model, the left column depicts genuine flickering between coexisting equilibria, while the right column shows a null trajectory where the noise amplitude is modulated to mimic flicker. Within each panel, the top trace is the state time series, the middle trace shows the classifier’s flicker (blue) and non-flicker (yellow) probabilities, and the bottom trace shows the moving variance. The red band marks a noise-driven interval used to probe false-alarm behavior in the null case; axes are aligned to enable direct visual comparison.	55

3.4	ROC performance on unseen dynamical systems. Receiver operating characteristic (ROC) curves comparing a variance-based indicator (orange) and the deep learning (DL) classifier (blue) for six nonlinear systems (A–F). Curves are based on 300 simulations per system under flickering and null conditions. Variance-based AUCs hover near 0.5, indicating poor discrimination when variance inflates in both cases, while DL-based curves are consistently bowed toward the upper-left, yielding markedly higher AUCs and capturing flickering-specific dynamics beyond variance growth.	59
3.5	Application of the deep learning classifier to empirical datasets. (A,B) Body temperature records from male and female edible dormice during the transition from active to hibernation states. (C) Standardised potassium concentrations from the Chew Bahir palaeoclimate record in southern Ethiopia, a proxy for aridity, spanning 9–3 kyr BP and capturing the termination of the African Humid Period. In each panel, the upper subfigure shows the original time series (Grey); the lower subfigure shows the deep learning ensemble probabilities for flickering (blue) and non-flickering (orange). Clear pre-transition flickering signals are detected in (A–C). (D,E) Two additional dormouse temperature records containing flickering that the classifier failed to detect, highlighting cases of missed detection. Probabilities are obtained by averaging outputs from models trained with sliding windows spanning 8%, 9.6%, 11%, 13%, 14%, and 16% of the total time series length.	62
4.1	Possibility of avoiding climate tipping point through social actions. In this figure, The Baseline model is illustrated in blue and the modified model with an additional tipping term is illustrated in orange. The dynamic of x (fraction of mitigators) versus time is also depicted in both plots in a smaller box. Two plots differ in the values of κ and R_{max} (GtC/year) that were used: (0.01, 5) for plot A, where it shows tipping behaviour, and (0.1, 1) for plot B, where it does not.	74
4.2	The difference in AUC versus social behaviour variables. In this figure, the difference in AUC is depicted against social behaviour parameters and R_{max} (GtC/year). The top plots represent the high-risk scenario ($T_c = 2$), and the bottom ones correspond to the low-risk scenario ($T_c = 3$). The plots illustrate the difference in AUC versus Social Learning Rate on the left, Net Cost of Mitigation in the middle, and Strength of Social Norm on the right.	76

4.3	Time to the tipping point for different emission and mitigation scenarios. Time to the tipping point is plotted against R_{max} (GtC/year) and κ for six different cases. The top plots represent the results of the high-risk scenario ($T_c = 2$), and the bottom ones show the results of the low-risk scenario ($T_c = 3$). The plots on the left, middle, and right depict the results for different values of the threshold ($d = 1.1, 1.25$, and 1.5).	77
4.4	Peak Temperature Anomaly for Different Emission and Mitigation Scenarios Peak temperature anomaly (Celsius) is depicted against R_{max} (GtC/year) and κ for the high-risk scenario (left) and the low-risk scenario (right). Isoclines show that as K increases, for the high-risk scenario, the decline in the peak of temperature is lower than low-risk scenario.	78
4.5	The difference in AUC versus critical temperature and social learning rate The Difference in Area under the curve (AUC) is depicted versus critical temperature and κ (left). Two examples representing two different choices of T_c and κ are chosen and illustrated on RHS. The plot at the top (orange circle) represents a case in which the modified model does not go through a tipping point, and the one at the bottom (blue star) represents a case of tipping point	79
4.6	The Coupled Tipping Dynamics Between Climate and Social Models As shown in panels A and B, for high values of delta, decreasing beta destabilises 0 and triggers the social model to tip to $x=1$. A more detailed illustration of this effect in panel C shows that beyond a specific threshold value of beta, the tipping term does not lead to a significant difference between the modified and baseline model, whereas below that threshold, the tipping element activates.	80
5.1	Illustration of the relative integrated absolute error (RIAE) metric. (A) Climate dynamics: comparison between spectrum and binary social models, with the shaded region indicating the integrated absolute difference after the onset of social dynamics (t_0) until the end of the evaluation window. (B) Social dynamics: comparison of mean opinion trajectories under spectrum and binary formulations, with shaded areas likewise representing the integrated L^1 gap.	90
5.2	Illustration of binary ($N = 2$) and spectrum ($N = 100$) replicator models coupled to the climate-carbon system. Panels show RIAE of temperature anomaly $T(t)$ (A, C, E, and G) and mean mitigation $\langle m(t) \rangle$ (B, D, F, and H)	93

5.3	Illustration of binary and spectrum FJ social models coupled to the climate-carbon system. Panels show RIAE of $T(t)$ (A, C, E, and G) and $\langle m(t) \rangle$ (B, D, F, and H) across sweeps of mean susceptibility λ_0 , runaway amplitude $R_{\max}$, adjustment rate Ψ , and interaction sensitivity A	96
5.4	Illustration of binary and spectrum replicator models coupled to the forest-grassland system. RIAE of forest fraction F (A, C, E, and G) and mean opinion $\langle m \rangle$ (B, D, F, and H) are shown for sweeps of v , κ , r_0 , and c , under weak ($h = 0.01$) and strong ($h = 0.1$) coupling.	98
5.5	Illustration of binary and spectrum FJ models coupled to the forest-grassland system. Panels report RIAE of forest fraction F (A, C, E, G, I, and K) and mean opinion $\langle X \rangle$ (B, D, F, H, J, and L) across sweeps of $R_{\max}$, Ψ , A , λ_0 , c , and v , under two ecological thresholds $F_c \in \{0.2, 0.8\}$	100
A.1	Deep learning architecture	127
A.2	(A–C) Synthetic null time series from three canonical bifurcation models with additive white noise all far from the bifurcation point: (A) Rosenzweig–MacArthur predator-prey system (B) Rosenzweig–MacArthur, (C) May’s harvesting model (D–F) Deep learning model predictions for the systems in (A–C), showing temporal evolution of bifurcation class probabilities. Probabilities are represented with distinguishable markers and line styles for print clarity: triangle (fold), square (transcritical), circle (Hopf), and x (null). In all subplots, the null class consistently exhibits the highest predicted probability compared to the other bifurcation types.	129
A.3	(A–C) Synthetic null time series from three canonical bifurcation models with additive white noise all far from the bifurcation point: (A) Van der pol system, (B) SIS model, (C) Neural Activation model (D–F) Deep learning model predictions for the systems in (A–C), showing temporal evolution of bifurcation class probabilities. Probabilities are represented with distinguishable markers and line styles for print clarity: triangle (fold), square (transcritical), circle (Hopf), and x (null). The null class consistently exhibits the highest predicted probability in all subplots compared to the other bifurcation types.	130
C.1	Sensitivity of climate impact to the social learning rate κ across different tipping strengths $R_{\max}$. Left: Signed derivative $\partial \text{AUC} / \partial \kappa$. Right: Absolute value $ \partial \text{AUC} / \partial \kappa $. Peak sensitivity occurs in an intermediate regime of κ , particularly when $R_{\max}$ is high.	135

C.2	Tornado plot showing the sensitivity analysis for the Difference in the area under the curve (left) and the peak of temperature (right). The yellow (green) bar shows the sensitivity values when the specified parameter has its upper (lower) bound value. The definitions of all parameters and variables are provided in a table in 5.2.	138
C.3	Historical record for Emission rate Vs Time is plotted. This data has been gathered from 1800 to 2017.	139

List of Tables

2.1	Model Performance Metrics Across Redness Levels	32
2.2	Confusion matrices for CNN–LSTM classifier performance across redness levels. Each matrix shows predicted (columns) versus true (rows) class counts.	33
2.3	Classification metrics (precision, recall, and F1-score) for the CNN–LSTM classifier under different redness values.	34
2.4	Model performance metrics with mean and standard deviation over multiple runs.	35
2.5	Comparison of overall performance metrics between CNN-LSTM and LSTM-only models.	37
D.1	Parameters of the replicator–climate–carbon coupled model (carbon and climate subsystem).	143
D.2	Parameters of the replicator–climate–carbon coupled model (social subsystem).	144
D.3	Parameters and numerical settings for the FJ–forest coupled model. Baselines are used unless a sweep over the stated range is performed.	145

Chapter 1

Introduction

This chapter provides a broad overview of the fundamental concepts that underpin the rest of the thesis. It is intended for a general audience and introduces the key ideas of complex systems, tipping points, dynamical systems, stochasticity, early warning signals, deep learning, and coupled socio-climate dynamics.

1.1 Complex Systems and Tipping Points

Complex systems are systems composed of many interacting components whose collective behaviour cannot be fully explained by examining the parts in isolation. Unlike simple mechanical systems, the interactions among the components are nonlinear, meaning that small changes in one part of the system can lead to disproportionately large effects in the whole [51, 64]. This nonlinearity gives rise to behaviours that are often counter-intuitive, such as sudden shifts, oscillations, or chaotic dynamics. Complex systems are therefore best understood not by reduction to their individual parts, but by studying the patterns and relationships that emerge from their interactions [182, 5]. A defining feature of complex systems is the presence of feedback loops. Positive feedback occurs when changes reinforce themselves, for example, melting ice reduces albedo, which in turn accelerates further warming. Negative feedback, by contrast, dampens deviations and promotes stability, for instance, predator–prey interactions can regulate population sizes. These feedbacks interact across different scales, often producing nonlinear thresholds or tipping points beyond which the system undergoes a qualitative transformation. The coexistence of stabilizing and destabilizing forces makes prediction and control especially challenging [146, 134]. Emergence is another hallmark of complex systems: the appearance of new properties at the system level that are not present in the individual components. Exam-

ples include the flocking of birds [205], market crashes in financial systems [191], or abrupt climate transitions [124]. In each case, individual elements follow simple rules, but their interactions generate collective outcomes that are unexpected and often difficult to reverse. Understanding emergence is therefore essential for anticipating critical transitions in natural and social systems, and it forms a central motivation for the approaches developed in this thesis [90]. Tipping points are thresholds in complex systems where a small change in external conditions can trigger a sudden and often irreversible shift in the system’s state. Before the threshold is reached, the system may appear stable and resilient, but once the tipping point is crossed, its dynamics reorganize qualitatively. Such transitions stand in stark contrast to gradual or linear change: while the forcing variable may increase smoothly, the system’s response is abrupt. The concept of tipping points is particularly important in disciplines where stability and resilience determine long-term outcomes, including climate science, ecology, economics, and public health [180]. Figure 1.1 depicts two real-world examples of empirical time series undergoing a tipping point.

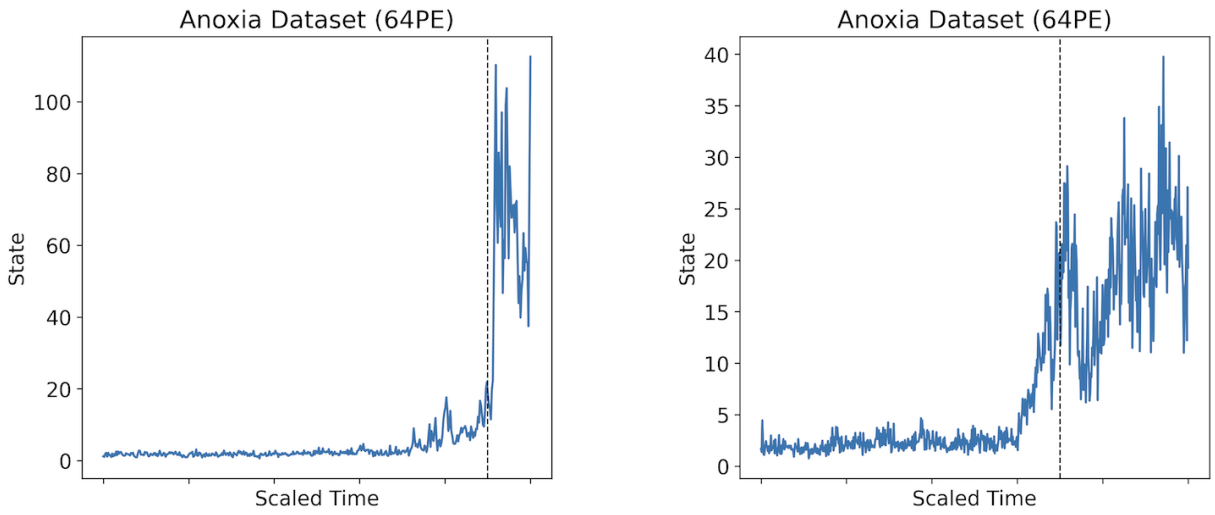


Figure 1.1: Empirical anoxia dataset (64PE) showing a tipping point, with the vertical dashed line indicating its approximate location.

Real-world examples highlight the importance of tipping points and critical transitions. In the climate system, processes such as the potential collapse of the Greenland or West Antarctic ice sheets [47], the dieback of the Amazon rainforest [86], or the weakening of the Atlantic Meridional Overturning Circulation [168] illustrate how small increases in temperature or emissions could provoke abrupt, global-scale consequences. In ecosystems,

shallow lakes can shift suddenly from clear to turbid states due to nutrient enrichment [179], coral reefs may collapse from coral to algal dominance after bleaching events [95], and forests can transform into grasslands when thresholds of fire frequency or drought stress are exceeded [86]. Financial markets also exhibit analogous dynamics: local disturbances such as bank failures or shifts in investor confidence can cascade through a tightly connected system, triggering sudden market crashes or global recessions. These examples underscore that tipping points are not abstract constructs but tangible risks that demand careful study and monitoring across disciplines[191].

1.2 Dynamical Systems and Bifurcation Theory

Dynamical systems provide a mathematical framework for describing how quantities evolve over time under a set of deterministic or stochastic rules. At their core, these systems consist of variables representing the state of interest, and equations, often differential or difference equations, that prescribe how the state changes in response to both internal interactions and external influences. This generality makes dynamical systems an essential tool across disciplines, from physics and engineering to biology, economics, and the social sciences. By formalising change in mathematical terms, they allow us to understand and predict patterns such as growth, decay, oscillations, and chaotic fluctuations[193]. A central focus in the study of dynamical systems is the identification of equilibrium states, or fixed points, where the system ceases to evolve unless perturbed. The stability of these equilibria determines whether small disturbances decay, leaving the system unchanged, or amplify, leading to a qualitative shift in behaviour. In continuous-time systems described by differential equations, this stability is often assessed using linearisation and eigenvalue analysis. In discrete-time systems, analogous criteria are applied to maps and iterated functions. These mathematical tools provide insight into how local dynamics can shape the global trajectories of the system [87]. Beyond equilibria, dynamical systems exhibit a rich variety of behaviours including limit cycles, quasi-periodic oscillations, and chaos. The study of these behaviours has revealed that even simple nonlinear systems can generate complex, unpredictable outcomes that nevertheless follow deterministic rules. This duality, order emerging from simplicity, and complexity arising from determinism, captures the essence of dynamical systems. It also explains why they are invaluable in exploring the conditions under which critical transitions occur, as well as in developing methods, such as those used in this thesis, for anticipating and interpreting tipping points in noisy, real-world systems[166, 133]. Bifurcation theory studies how the qualitative behaviour of a dynamical system changes as parameters vary. Broadly, bifurcations can be classified as local or global, depending on whether the transition is governed by changes in the stability

of equilibria in the immediate neighbourhood of a fixed point, or by large-scale structural reorganisations of the system’s trajectories. Distinguishing between these two classes is important, because they give rise to different types of critical transitions and early warning signatures[113, 110]. Local bifurcations occur when small parameter changes cause an equilibrium to lose or gain stability, often through the crossing of an eigenvalue across the imaginary axis. Classic examples include the fold (or saddle-node) bifurcation, where two equilibria collide and annihilate each other; the transcritical bifurcation, where two equilibria exchange stability; and the Hopf bifurcation, where a fixed point gives rise to a limit cycle. These transitions are “local” because they can be analysed through linearisation and centre manifold reduction near the equilibrium point, making them mathematically tractable and widely studied in applications[36]. Figure 1.2 depicts three simulated stochastic time series generated from three mathematical models exhibiting transcritical, Hopf, and fold bifurcations.

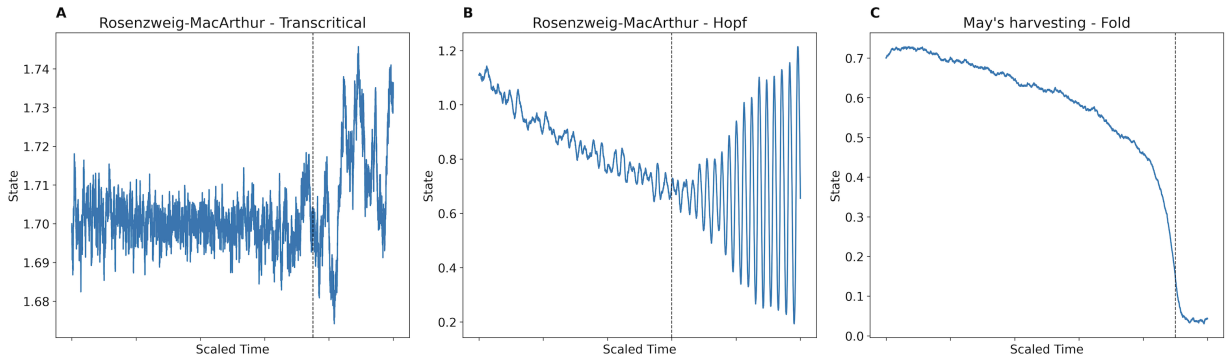


Figure 1.2: Bifurcation dynamics in three canonical models: Rosenzweig–MacArthur with a transcritical bifurcation (left), Rosenzweig–MacArthur with a Hopf bifurcation (middle), and May’s harvesting model with a fold bifurcation (right). Vertical dashed lines indicate the approximate locations of the transitions.

Global bifurcations, by contrast, involve qualitative changes in the system’s trajectories that cannot be captured solely by examining equilibria in their immediate neighbourhood. They often arise when invariant manifolds or attractors interact on a larger scale. Examples include homoclinic bifurcations, where trajectories loop back to the same saddle point, and heteroclinic bifurcations, where trajectories connect different saddles. These events can trigger dramatic changes such as the sudden onset of chaos or large-amplitude oscillations. Because global bifurcations depend on the structure of the system’s phase space as a whole, they are harder to predict and often lack simple early warning signals. Nonetheless, they play a critical role in many natural and social systems, highlighting the need for

approaches, such as those developed in this thesis, that go beyond local stability analysis [77, 113]. Nonlinear dynamical systems can often appear overwhelmingly complex due to the large number of terms in their governing equations. However, near a bifurcation point, much of this complexity is unnecessary for describing the essential behaviour of the system. Normal form theory provides a systematic way to reduce nonlinear equations to their simplest equivalent representation while preserving the qualitative dynamics of interest. Through a series of coordinate transformations, higher-order terms that do not affect the local behaviour are eliminated, leaving behind a canonical equation, the normal form, that captures the dynamics near the bifurcation [113, 8]. The following are the normal form equations of the Fold, Transcritical, and Hopf bifurcations, which are frequently employed in this thesis.

$$\begin{aligned}
\textbf{Fold:} \quad & \dot{x} = \mu - x^2, \\
\textbf{Transcritical:} \quad & \dot{x} = \mu x - x^2, \\
\textbf{Hopf:} \quad & \begin{cases} \dot{x} = \mu x - \omega y - a(x^2 + y^2) x, \\ \dot{y} = \omega x + \mu y - a(x^2 + y^2) y, \end{cases}
\end{aligned} \tag{1.1}$$

The power of normal forms lies in their universality. Many different systems, when approaching the same type of bifurcation, reduce to the same normal form despite having very different underlying mechanisms. For example, a fold bifurcation in an ecological population model and in a mechanical oscillator can both be reduced to the same simple quadratic normal form. This universality allows researchers to study a small set of canonical equations rather than the infinite diversity of real-world systems, greatly simplifying both theoretical analysis and computational modelling [113, 8]. In practice, normal forms enable clearer understanding of early warning signals and the conditions under which tipping points occur. By stripping a system down to its core dynamical ingredients, they allow analytical results, such as the scaling of variance, autocorrelation, or recovery rates near bifurcations, to be derived in a transparent way. In this thesis, normal forms serve not only as the theoretical foundation for constructing training datasets for deep learning models, but also as a bridge between mathematical theory and empirical application. They demonstrate how rigorous simplification can provide both insight and practical tools for anticipating critical transitions in complex, noisy systems. As a system approaches a tipping point, its resilience diminishes, and this weakening often leaves detectable statistical fingerprints in the time series of system variables. One of the most prominent phenomena is known as critical slowing down: the progressive decline in recovery rate following small perturbations. When a system is close to a bifurcation, the restoring forces that normally return it to equilibrium become weaker, so deviations persist longer. This manifests as a

slowdown in the system's ability to recover, and it can be identified in empirical data even when the underlying dynamics are not explicitly known [181, 43]. Figure 1.3 illustrates how the basin of attraction in a dynamical system changes as it approaches a bifurcation, leading to an increase in the autocorrelation of the corresponding stochastic time series.

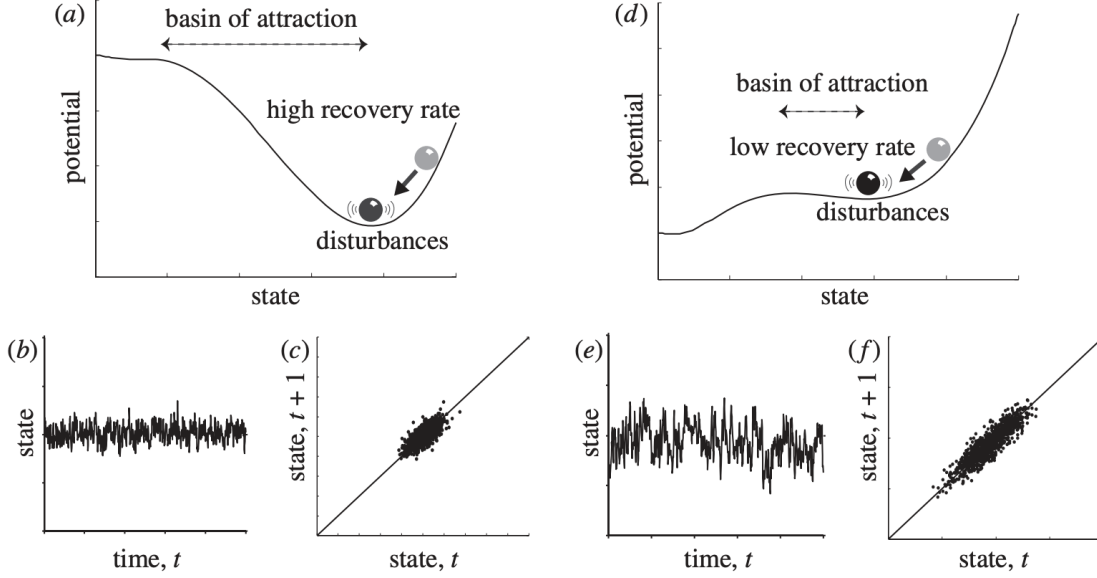


Figure 1.3: Loss of resilience near a bifurcation: as the system approaches the tipping point, recovery from perturbations slows and autocorrelation increases. Image taken from [125]

Critical slowing down is associated with measurable statistical indicators that serve as early warning signals. Increased variance is one such indicator: as the system becomes less stable, fluctuations around equilibrium grow larger because perturbations are no longer quickly dampened. Another widely used signal is elevated lag-1 autocorrelation, which reflects the persistence of successive deviations over time. In other words, as the system approaches a critical transition, it becomes increasingly likely to remain high (or low) in the next time step if it was high (or low) in the previous one. Mathematically, the lag-1 autocorrelation of a time series $\{x_t\}$ is defined as

$$\rho_1 = \frac{\text{E}[(x_t - \mu)(x_{t-1} - \mu)]}{\sigma^2},$$

where μ and σ^2 denote the mean and variance of the process, respectively. A rise in ρ_1 indicates increasing memory and slower recovery in the system's dynamics. Together,

variance and autocorrelation provide complementary evidence of reduced stability and approaching critical transitions [35, 215].

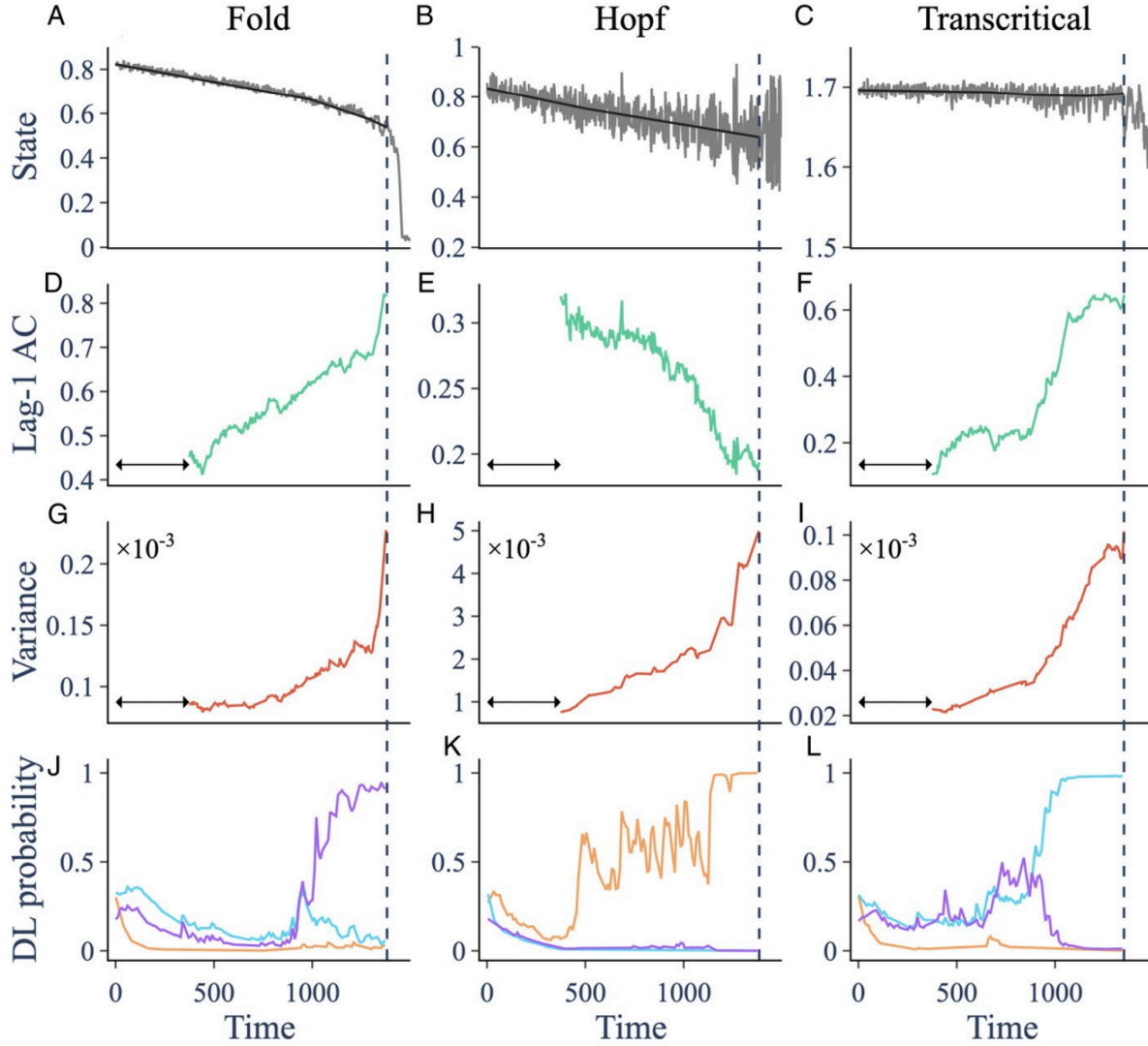


Figure 1.4: Change in variance and lag-1 autocorrelation in fold, transcritical, and Hopf systems, illustrating their role as early warning signals of bifurcations. Image taken from [34]

Although these indicators have been applied successfully in a variety of contexts, includ-

ing ecology, climate science, and epidemiology, they also have limitations. Variance and autocorrelation can be influenced by external drivers, changes in measurement resolution, or noise characteristics, sometimes producing false alarms or missing true signals. Moreover, they typically indicate that a transition is near, but not what kind of bifurcation is approaching. For this reason, researchers have sought to augment traditional early warning signals with new approaches, such as machine learning and deep learning methods, which can capture more complex patterns of instability. This thesis builds upon the foundations of classical EWS while exploring how modern data-driven methods can extend their power in detecting and classifying critical transitions[32, 34].

1.3 Stochasticity and Noise

White noise refers to a stochastic process with no temporal correlation: each fluctuation is independent of the previous one, and the power spectrum is flat across all frequencies. Because of its mathematical simplicity, white noise has been widely used in modelling as a default representation of random perturbations. It provides a convenient baseline for studying how dynamical systems respond to variability, since its lack of memory ensures that fluctuations are uncorrelated over time[2]. In contrast, many real-world systems are more accurately described by coloured noise, which exhibits temporal correlation and frequency-dependent structure. This arises because environmental and internal fluctuations often have finite correlation times rather than being independent between successive steps. As a result, the noise spectrum deviates from the flat case and may exhibit dominance of low-frequency components or intermediate-frequency structure, thereby influencing the apparent stability and early-warning indicators of the underlying dynamics. Red noise, for example, has stronger low-frequency components and is highly persistent, while pink noise lies between white and red, with power spectral density inversely proportional to frequency. Other types include blue and violet noise, which emphasize higher frequencies. These different noise structures can profoundly influence system dynamics by shaping the persistence and amplitude of perturbations [2, 203]. The distinction between white and coloured noise is particularly important in the study of critical transitions. Red noise, in particular, can mask or mimic the statistical signatures of critical slowing down, complicating efforts to detect approaching tipping points. This occurs because red noise itself exhibits strong temporal autocorrelation and elevated low-frequency variance features that are also expected from a system losing stability. Consequently, increases in variance or lag-1 autocorrelation may arise from the coloured structure of the external forcing rather than from intrinsic dynamical slowing down. In such cases, the apparent early-warning signals can be false positives, reflecting persistent stochastic forcing rather than genuine

proximity to a bifurcation. Disentangling these effects therefore requires careful spectral analysis or the use of models that explicitly account for noise color and system memory. Recognising the type of noise present is therefore essential for both theoretical modelling and empirical analysis [42]. Noise is not merely a mathematical abstraction but an inherent feature of natural and social systems. In climate dynamics, random fluctuations arise from weather variability, volcanic eruptions, or solar forcing [79]. In ecosystems, demographic stochasticity, migration, and environmental variability introduce randomness into population dynamics [118]. Financial systems exhibit noise through unpredictable market shocks, investor sentiment, and sudden changes in liquidity [139]. Across these examples, noise is inseparable from the underlying deterministic forces. The ubiquity of noise reflects the fact that no system exists in isolation. Each is embedded within a larger context that continually perturbs it, often in ways that are challenging to represent or control with precision. Even highly engineered systems, such as power grids or communication networks, face random disturbances due to hardware failure, external shocks, or cyberattacks. Far from being negligible, these stochastic influences can accumulate over time or interact with deterministic dynamics to produce large-scale outcomes [54]. For scientists and policymakers, this ubiquity means that predictive models must account for uncertainty as an intrinsic property rather than a nuisance. Ignoring stochasticity can lead to overconfident forecasts or failure to anticipate extreme events. By treating noise as a central feature of system dynamics, researchers can develop more realistic models and more robust strategies for managing risk. Noise does not simply obscure system dynamics; it can actively shape them. In some cases, random fluctuations destabilise equilibria and drive systems across thresholds earlier than would be expected under deterministic forcing alone. This phenomenon, sometimes called noise-induced tipping, highlights how randomness can trigger transitions even when control parameters are far from bifurcation values. Such effects are particularly important in climate and ecological systems, where noise can push the system into undesirable states [72]. Conversely, noise can also have stabilising or constructive effects. In stochastic resonance, for instance, a moderate level of noise enhances the system's ability to respond to weak periodic signals, effectively amplifying otherwise undetectable dynamics. Similarly, in some ecological and neuronal systems, noise helps maintain diversity or synchrony [212, 197]. These examples demonstrate that randomness is not always destructive; under certain conditions, it can play a functional role in sustaining complex behaviour [20, 143]. The interaction between noise and stability is especially relevant to the study of early warning signals. Coloured noise can mask indicators such as rising autocorrelation or variance, or generate spurious patterns that mimic them. At the same time, understanding the stochastic environment allows researchers to refine detection methods and design more robust predictive tools. In this thesis, special attention is given to how different noise structures interact with bifurcations and tipping points, both in theoretical

models and in empirical applications [26].

1.4 Early Warning Signals (EWS)

Traditional early warning signals are based on simple statistical properties of time series that change systematically as a system approaches a tipping point. Two of the most widely used indicators are variance and autocorrelation at lag one. As resilience declines, perturbations are less strongly damped, leading to larger fluctuations in state variables and greater similarity between successive observations. These patterns reflect the phenomenon of critical slowing down and have been applied successfully in ecological, climatic, and epidemiological studies [215, 35]. Additional indicators include skewness and kurtosis, which capture shifts in the distribution of fluctuations, and spectral properties that measure changes in the relative importance of low- versus high-frequency variability. In spatial systems, indicators such as spatial variance and correlation length provide analogous signatures of approaching transitions. Collectively, these methods offer a relatively simple and interpretable toolkit for detecting reduced stability in both models and empirical data [78]. One of the strengths of traditional indicators is their broad applicability across disciplines. Because they require only time series data and not a detailed model of the system, they can be used in settings where mechanisms are poorly understood. This generality explains why variance and autocorrelation have become benchmarks for studying critical transitions and why they remain an important reference point for more advanced methods. Despite their usefulness, traditional statistical indicators face significant limitations. Many of the same signals that indicate reduced resilience, such as rising variance or autocorrelation, can also be produced by external drivers, measurement noise, or slow parameter drift unrelated to an impending tipping point. As a result, these methods can generate false alarms or miss true warnings, reducing their reliability in practice [34, 32]. Another limitation is that traditional EWS typically do not reveal the type of bifurcation that the system is approaching. While they may signal that resilience is declining, they cannot easily distinguish between fold, transcritical, or Hopf bifurcations, each of which has different implications for system behaviour. For policymakers or managers who need actionable information, the inability to classify the nature of the transition restricts the practical utility of these indicators [32, 34]. Furthermore, many indicators assume stationarity and rely on long, high-quality time series. In reality, empirical datasets are often short, irregularly sampled, and influenced by confounding processes. These conditions can erode the statistical power of EWS and make it difficult to establish robust thresholds for detection. These challenges motivate the exploration of complementary approaches that can handle complexity and noise more flexibly. Given the limitations of classical indicators, researchers have turned to machine

learning and deep learning as promising alternatives for detecting early warning signals [34]. Unlike traditional methods, which rely on a small set of predefined statistics, deep learning models can learn relevant features directly from the data [119]. Convolutional neural networks (CNNs) can extract local temporal patterns, while recurrent neural networks (RNNs) and long short-term memory (LSTM) networks capture long-range dependencies, making them well-suited for analysing time series that contain complex, nonlinear signatures of instability [120, 88]. Deep learning methods offer the possibility of classifying not only the presence of an impending tipping point but also the specific type of bifurcation. By training on synthetic datasets constructed from canonical models, neural networks can learn to recognise fingerprints of fold, transcritical, and Hopf bifurcations, as well as distinguish them from null cases. This capacity goes beyond what variance and autocorrelation can provide, potentially yielding richer and more actionable predictions [34, 58]. Another key motivation is robustness. Deep learning models can generalise across diverse systems and noise conditions, and they are less dependent on long, stationary time series. They also open the door to integrating empirical datasets with theoretical simulations, building hybrid frameworks that combine mechanistic insight with data-driven flexibility. These advantages form the foundation for the deep learning approaches developed in this thesis, which seek to extend the reach of early warning signals into settings where traditional indicators are insufficient. Nevertheless, deep learning methods also face important limitations. They typically require large and representative training datasets, and their performance can be sensitive to biases or imbalances in the training data. Moreover, the internal representations learned by such models are often opaque, making it difficult to interpret their decisions in mechanistic or causal terms. These challenges underscore the importance of combining deep learning with theory-based approaches, careful validation, and transparent model interpretation.

1.5 Deep Learning for Dynamical Systems

Artificial intelligence (AI), and in particular deep learning, has emerged as a powerful framework for analysing data generated by nonlinear, noisy systems. Traditional mathematical models often require strong assumptions about system structure, which can be limiting when dynamics are complex, high-dimensional, or poorly understood. By contrast, Deep learning models can extract patterns directly from data, identifying patterns and dependencies that are difficult to specify analytically. This makes it especially useful for systems near tipping points, where nonlinear feedbacks and stochastic fluctuations interact in complicated ways. Another advantage of deep learning is its capacity to capture subtle, multiscale features. Complex systems often exhibit behaviours that unfold across

different temporal or spatial scales, for example, long-term trends interspersed with sudden bursts of activity. Conventional statistical methods may overlook such patterns, but neural networks can automatically detect them through hierarchical representations [34, 58]. This ability to encode both local and global features is particularly relevant when studying bifurcations, critical slowing down, and flickering [34, 58]. Deep learning has proven particularly effective in bifurcation detection, where models trained on synthetic simulations can generalize to empirical datasets. By leveraging trajectories generated from canonical dynamical systems, such models can learn to recognise universal precursors of instability patterns that transcend specific equations or parameterisations. These trained models can then be applied to empirical data, offering a way to anticipate transitions even when the underlying equations are unknown. Such robustness

Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) are designed to detect and hierarchically represent local patterns in structured data such as images, signals, or time series. They are composed of several stages convolution, pooling, and fully connected layers that transform raw input data into abstract representations useful for classification or regression tasks. Figure 1.5 illustrates the main components and information flow within a CNN. The core operation in a CNN is the **convolution**, in which a small kernel (or filter) slides across the input data and computes weighted sums over local neighbourhoods. Mathematically, a one-dimensional convolutional layer computes

$$y_i = \sigma \left(\sum_j (w_j * x_j)_i + b \right),$$

where x_j denotes the input signal or feature map, w_j the convolutional kernel, b a bias term, and $\sigma(\cdot)$ a nonlinear activation function such as ReLU or tanh. Each filter is trained to respond to a specific type of pattern, for instance, an oscillation, spike, or sudden change in amplitude. In time-series applications, these filters act as feature detectors that identify short-term structures associated with dynamical behaviour, such as transient bursts or fluctuations preceding a bifurcation. Following convolution, the network applies a nonlinear activation function to introduce flexibility and allow the model to approximate nonlinear mappings. A pooling layer then down-samples the resulting feature maps by summarising information within small regions, typically using a maximum or average operation. Pooling serves two purposes: it reduces the computational complexity of subsequent layers and provides translational invariance, meaning that small shifts in the input signal do not alter the extracted features.

After several rounds of convolution and pooling, the resulting feature maps are flattened into a single vector and passed through one or more fully connected layers. These layers integrate information across all detected features, enabling the network to infer global relationships. In classification tasks, the final layer typically uses a softmax activation function to produce a probabilistic distribution over output categories. In the context of dynamical systems, this architecture allows CNNs to capture characteristic motifs such as oscillatory precursors, abrupt transitions, or shifts in variance that can signal proximity to bifurcations [120, 119]. By learning these features directly from data, CNNs can identify early-warning patterns even when explicit analytical indicators are unavailable.

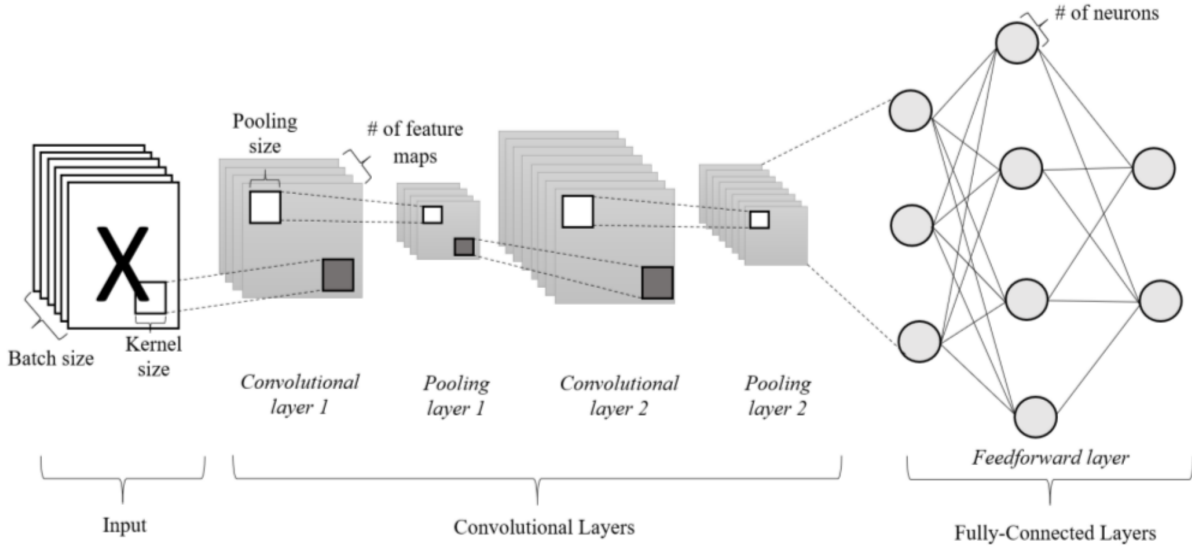


Figure 1.5: Illustration of a convolutional neural network (CNN). Convolutional layers extract local features through filtering operations, pooling layers reduce dimensionality, and fully connected layers integrate the learned representations to perform classification or prediction. The image is taken from [6].

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) Networks

While CNNs are effective at identifying local patterns, they treat input segments largely independently. Recurrent neural networks (RNNs) address the complementary challenge of modelling temporal dependencies. In an RNN, information from previous time steps is

stored in a hidden state h_t that evolves as

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b),$$

where W_h and W_x are weight matrices and b is a bias vector. This recursive structure enables the network to retain memory of past dynamics, making it suitable for systems where the current state depends on prior evolution. However, classical RNNs often suffer from vanishing or exploding gradients during training, which limits their performance on long sequences [62].

To overcome these limitations, *long short-term memory* (LSTM) networks were developed [88]. LSTMs extend the RNN architecture by introducing a cell state c_t , which acts as an explicit memory pathway through time. The flow of information into, within, and out of this memory cell is regulated by three multiplicative gates, each controlled by a sigmoid activation function and one candidate update term. The key operations are as follows:

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) && \text{(forget gate),} \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) && \text{(input gate),} \\ \tilde{c}_t &= \tanh(W_c[h_{t-1}, x_t] + b_c) && \text{(candidate cell update),} \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t && \text{(cell state update),} \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) && \text{(output gate),} \\ h_t &= o_t \odot \tanh(c_t) && \text{(hidden state update),} \end{aligned}$$

where $\odot$ denotes elementwise multiplication, and each W_* and b_* represent trainable parameters associated with their respective gates. The mechanism is illustrated in Figure 1.6. The forget gate f_t controls which parts of the previous cell state c_{t-1} are retained. The input gate i_t and candidate update $\tilde{c}_t$ determine which new information is incorporated. The output gate o_t regulates how much of the internal memory c_t influences the current hidden state h_t . This gating architecture enables LSTMs to preserve relevant information over long temporal horizons while filtering out noise and transient fluctuations. Panel (a) of Figure 1.6 depicts the internal structure of an LSTM cell, showing the interaction between the three gates and the memory cell.

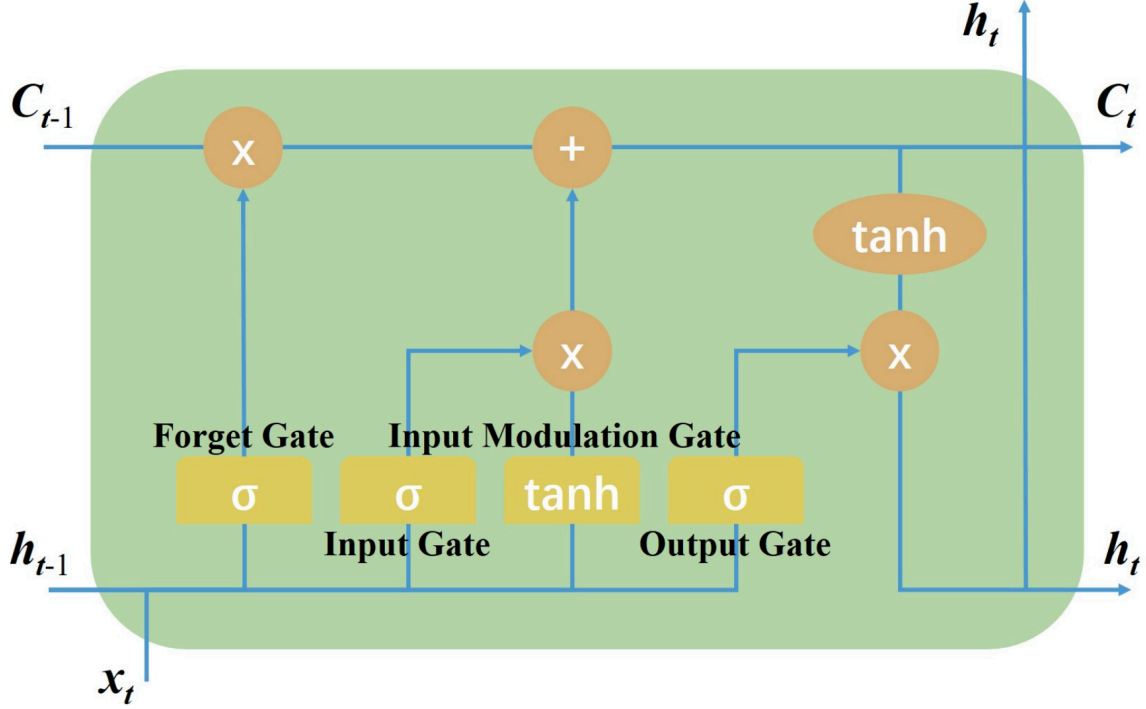


Figure 1.6: Illustration of a long short-term memory (LSTM) network, highlighting its memory cells and gating mechanisms for capturing long-range temporal dependencies. The image is taken from [107].

Example: CNN–LSTM hybrid for bifurcation detection

A CNN–LSTM hybrid model can be trained on synthetic time series generated from canonical normal-form systems, such as fold, Hopf, and transcritical bifurcations [34]. The convolutional layers first extract short-term temporal motifs, including oscillatory bursts or asymmetric fluctuations preceding critical transitions. The LSTM layers then integrate these local representations across time, learning the long-range dependencies that signal the system’s approach to instability. This architecture mirrors the analytical process of identifying early-warning indicators variance, autocorrelation, and skewness but learns these relationships directly from data rather than through explicit mathematical derivation.

Deep learning architectures originally gained prominence in fields such as image and speech recognition. In speech recognition, CNNs and LSTMs are used together to extract acoustic features and model sequential phoneme patterns, enabling robust transcription

of spoken language. These successes illustrate the power of neural networks to capture patterns in noisy, nonlinear signals that vary across multiple timescales [109, 119]. In finance, deep learning has been applied to forecasting asset prices, modelling volatility, and managing risk. Market data are notoriously noisy and influenced by nonlinear feedbacks, yet neural networks can uncover predictive signals that elude traditional econometric models. For example, LSTM networks have been used to forecast exchange-rate dynamics and credit risk, while CNNs detect patterns in high-frequency trading data [81, 70]. In climate science, deep learning has been applied to predict extreme events, downscale climate projections, and detect early-warning signals of tipping points. Neural networks can integrate large, heterogeneous datasets—satellite imagery, paleoclimate records, and model simulations and extract relevant dynamical features directly from them [38, 34].

Taken together, CNNs and RNNs provide a bridge between data-driven pattern recognition and dynamical-systems analysis. Their layered architecture functions analogously to constructing successive approximations of a system’s vector field, where each layer refines the representation of the evolving state. This dual capability, capturing both mechanistic and empirical regularities, makes deep learning a powerful framework for bifurcation detection and early-warning signal analysis.

1.6 Flickering as a Precursor of Collapse

Flickering refers to the phenomenon in which a system, under stochastic forcing, repeatedly and intermittently switches between alternative stable states before an irreversible transition occurs. It reflects noise-induced excursions across potential barriers separating coexisting attractors, leading to abrupt jumps in state variables rather than smooth relaxation near equilibrium. Although the term “flickering” is typically used in this stochastic context, similar intermittent switching can also arise deterministically, for example, through chaotic or crisis-induced intermittency when the system’s intrinsic dynamics generate alternating behaviour between distinct regimes [44]. It is therefore important to distinguish stochastic flickering from deterministic intermittency, while recognising that both represent forms of transient switching that may precede critical transitions. As such, flickering is a hallmark of bistable or multistable systems, where multiple attractors coexist and random perturbations can temporarily push the system from one basin of attraction into another [45, 207]. Mathematically, flickering emerges in systems close to a bifurcation when the depth of potential wells becomes shallow. Perturbations that were once too small to induce a state shift can now cross reduced stability barriers, producing transient episodes in which the system explores alternative states (Figure 1.7). These excursions are not yet permanent, but they foreshadow the instability of the current equilibrium [45]. Be-

cause flickering is inherently noise-driven, it can occur even in the absence of deterministic parameter changes.

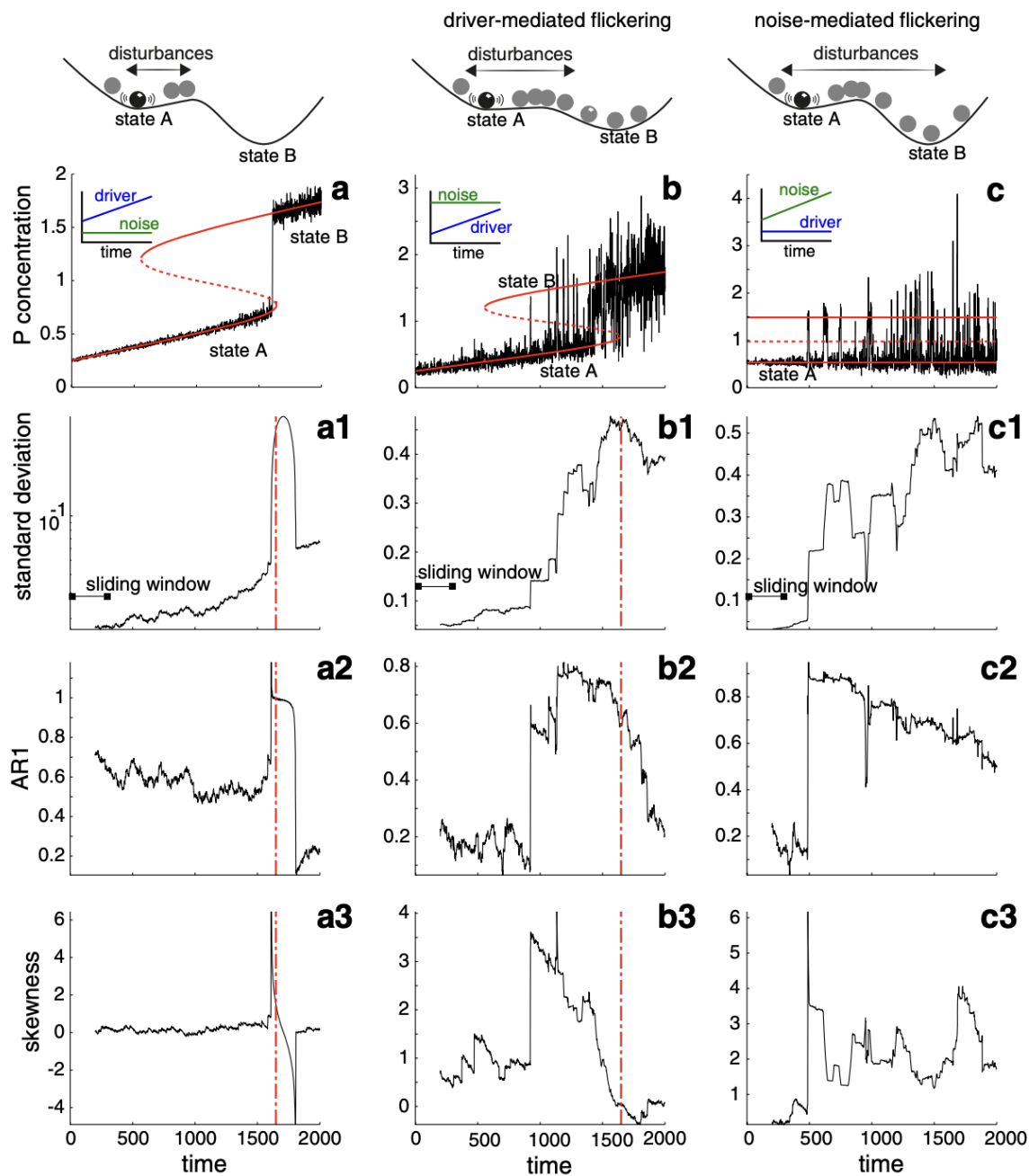
Detecting flickering therefore requires analytical or computational methods capable of distinguishing noise-induced switching from random fluctuations that mimic flickering-like variability within a single state [45]. Flickering provides valuable information about the resilience of a system. When resilience is high, the system remains confined to a single attractor, and noise-induced shifts are rare. As resilience declines, however, the boundaries between basins of attraction weaken, allowing flickering to occur more frequently.

Figure 1.7 illustrates three characteristic pathways toward a critical transition. In panel (a), the system exhibits **classical critical slowing down**: as the control parameter changes gradually, recovery from perturbations becomes slower, leading to increasing variance and autocorrelation before the transition. In panels (b) and (c), the system instead displays **flickering**, characterised by repeated excursions between coexisting states. Panel (b) shows *driver-mediated flickering*, where slow changes in the underlying driver bring the system intermittently across the stability boundary, while panel (c) illustrates *noise-mediated flickering*, where stochastic forcing alone triggers transitions between alternative attractors even without parameter drift. The corresponding time-series indicators (rows a1–a3, b1–b3, and c1–c3) show how standard deviation, lag-1 autocorrelation (AR1), and skewness evolve over time. Both types of flickering lead to strong increases in variance and skewness, reflecting the system’s intermittent switching, but their dynamical origins differ. Together, these cases highlight how flickering complements critical slowing down as a stochastic early-warning signal of diminishing resilience.

Figure 1.7: Comparison of dynamical regimes leading to transitions. (a) Classical critical slowing down: as the system approaches a bifurcation, recovery from perturbations slows, increasing variance and autocorrelation. (b) Driver-mediated flickering: slow changes in external forcing cause repeated transitions between coexisting states. (c) Noise-mediated flickering: random perturbations alone trigger state switching even without parameter drift. Corresponding panels (a1–c3) show the evolution of variance, lag-1 autocorrelation (AR1), and skewness computed over sliding windows. Image adapted from [45].

slowing down prior to transition

flickering prior to transition



The detection of flickering is particularly useful in systems where critical slowing down is difficult to observe or is confounded by external noise. Because flickering arises from stochastic exploration of the state space, it complements variance and correlation-based signals by providing a different line of evidence for declining resilience. Integrating these signals improves confidence in early warning frameworks [45]. Flickering has been documented in numerous empirical systems, highlighting its broad relevance. In palaeoclimate records, for example, sediment cores from the African Humid Period and other abrupt climate transitions show evidence of repeated switches between wet and dry states before permanent desertification. These fluctuations are interpreted as noise-driven flickering that preceded large-scale climate regime shifts [199]. Similar dynamics have been reported in coral reef ecosystems, where temporary shifts between coral and algal dominance anticipate long-term degradation [207]. Flickering has also been detected in physiological systems, including body temperature records of hibernating animals. For instance, dormice show intermittent switching between active and torpid states before entering prolonged hibernation [156]. These diverse examples demonstrate that flickering is not limited to theoretical models but is a real and measurable phenomenon across disciplines, reinforcing its role as a precursor of collapse [156].

Coupled socio–climate dynamics

The relationship between human societies and the climate system is deeply reciprocal. On the one hand, human activity drives climate change primarily through greenhouse gas emissions from fossil fuel use, land-use change, and industrial processes. On the other hand, the impacts of climate change such as rising temperatures, altered precipitation patterns, and increased frequency of extreme events feed back into human systems by shaping economic activity, migration, political stability, and public health. This bidirectional coupling means that neither the climate system nor social dynamics can be fully understood in isolation [124, 93].

Figure 1.8 illustrates this two-way coupling between the climate and social systems. The diagram highlights how human actions influence the climate system through emissions, and how climate impacts, in turn, shape social responses. Specifically, greenhouse gas emissions from the social system enter the atmosphere, altering the climate state through changes in temperature and precipitation. These climatic changes generate regional impacts such as droughts, floods, or agricultural losses that directly affect societies. Social perception of these impacts then feeds into behavioral models of decision-making, which determine future mitigation and adaptation efforts. Through this feedback loop, perceptions and behaviors within the social system influence emissions trajectories, which subsequently modify the climate system again.

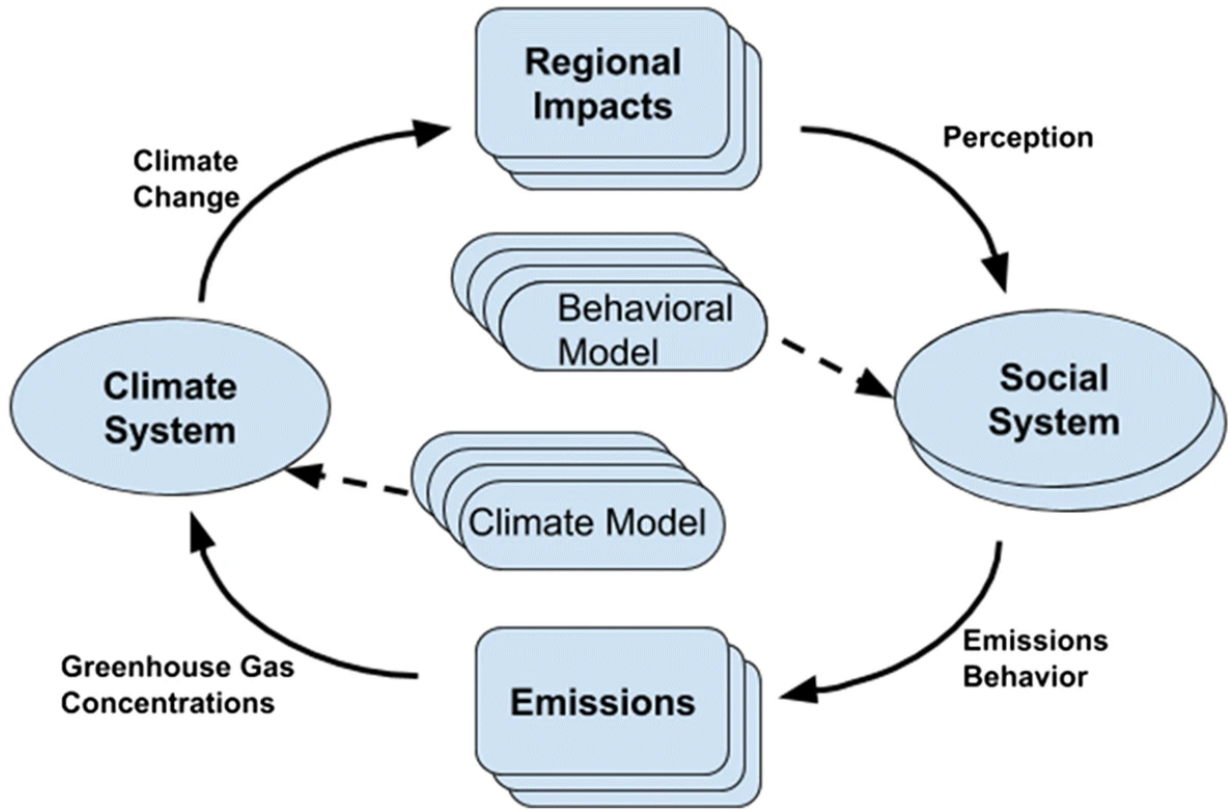


Figure 1.8: Schematic diagram of feedbacks between the social and climate systems, mediated by emissions, regional impacts, and behavioral responses. Image adapted from [18].

From a modelling perspective, coupling human behaviour with climate dynamics introduces additional layers of complexity. Decisions about mitigation, adaptation, and resource allocation are not purely rational but are influenced by cultural norms, economic incentives, and social interactions. These decisions collectively determine emission pathways, which then influence the trajectory of the climate system. Understanding these feedbacks requires bridging disciplines and combining tools from dynamical systems, ecological modelling, and social science [157].

Importantly, the socio–climate feedback loop can amplify risks. For example, delayed mitigation efforts allow greenhouse gas concentrations to rise, pushing the climate system closer to dangerous tipping points [30]. Conversely, visible climate impacts can accelerate public demand for action, creating social momentum for decarbonisation. The dynamic interplay between human choices and climatic responses therefore lies at the heart of pre-

dicting long-term outcomes and informing effective policy [14].

Just as natural systems can experience tipping points, social systems are also subject to sudden and nonlinear change. A **social tipping point** occurs when small shifts in behaviour, policy, or perception cascade through a population, leading to widespread adoption of new norms or technologies. Examples include the rapid uptake of renewable energy, sudden changes in public opinion on climate policy, or the decline of once-dominant industries [99, 60, 98]. These social transitions often appear abrupt from the outside but emerge from cumulative interactions among individuals and institutions [183, 37].

The significance of social tipping points lies in their ability to accelerate or delay climate transitions. Positive tipping points such as the rapid diffusion of clean technologies can dramatically reduce emissions and help avoid dangerous climate thresholds. Negative tipping points such as the erosion of public trust in institutions or the spread of misinformation can hinder collective action and exacerbate risks. Understanding the conditions that trigger social tipping dynamics is therefore critical for anticipating pathways toward either sustainability or collapse [157].

Moreover, social tipping points offer a potential source of hope in the context of climate change. While physical tipping points are often associated with irreversible losses, social systems can undergo rapid transformations that create opportunities for mitigation and resilience. By leveraging social dynamics through policy interventions, communication strategies, or grassroots movements—it may be possible to steer societies toward more sustainable futures even when climate risks are severe.

To represent how individuals make choices about climate mitigation, researchers often use simplified models of social opinion dynamics. Binary models assume that individuals adopt one of two discrete positions, such as supporting or opposing mitigation. These models are analytically tractable and can capture broad trends in how opinions spread through imitation, social learning, or normative pressure. Despite their simplicity, binary formulations often reproduce large-scale patterns of collective behaviour observed in empirical settings [30].

Spectrum models, by contrast, allow individuals to hold a range of positions along a continuum. This richer representation captures the diversity of opinions in real societies, where attitudes toward mitigation vary from strong opposition to enthusiastic support. Spectrum models can reveal subtler dynamics, such as gradual opinion shifts, polarisation, and the coexistence of multiple stable opinion distributions. They also highlight the role of heterogeneity in shaping collective responses to climate challenges [48, 169].

Comparing binary and spectrum models is crucial for understanding socio-climate interactions. While spectrum models offer greater descriptive richness, binary models remain far more common because they are simpler to analyse, easier to parameterise, and computationally tractable. A key question addressed in this thesis is whether binary models are

sufficient to capture the essential features of coupled socio–climate dynamics or whether spectrum models provide fundamentally different insights. This comparison clarifies the trade-offs between simplicity and realism in modelling human behaviour within climate systems.

1.7 Why These Topics Matter

The study of tipping points, bifurcations, and early warning signals is motivated by the urgent need to anticipate critical transitions before they occur. Once a system crosses a tipping point, recovery is often slow, costly, or even impossible. By developing methods to detect warning signals in advance, researchers aim to provide the opportunity for timely interventions that can prevent catastrophic outcomes. Prediction is not only about forecasting the exact timing of collapse but also about identifying conditions under which systems are at risk. Early warning indicators can highlight vulnerable states and reveal whether resilience is declining, even if the precise moment of transition cannot be determined. This shift from reactive to proactive management represents a critical change in how society approaches complex risks. Ultimately, the ability to predict and prevent collapse is about safeguarding the future of human and natural systems. Whether through stabilising ecosystems, preserving climate equilibrium, or maintaining financial stability, the tools developed in this field provide a foundation for informed decision-making. By anticipating where and when critical transitions might occur, we can allocate resources more effectively and design adaptive strategies that minimise the risks of irreversible change. The topics explored in this thesis are directly relevant to the policy challenges of our time. Climate policy, for example, must contend with nonlinear risks that are difficult to capture using traditional linear models. Recognising the possibility of tipping points changes the cost–benefit calculus of mitigation: the risks of inaction become far greater once the potential for abrupt and irreversible transitions is considered. Early warning frameworks could provide input for international agreements and national climate strategies. In the domain of risk management, the ability to detect declining resilience can improve preparedness across sectors. Financial institutions seek to avoid systemic crises, public health agencies monitor for epidemic outbreaks, and governments plan for natural disasters. In each case, understanding the dynamics of critical transitions allows decision-makers to design more resilient systems and respond more effectively to emerging threats. Societal resilience more broadly depends on anticipating shocks and adapting before they cause irreversible harm. By integrating insights from dynamical systems, stochastic processes, and machine learning, we can better understand how societies interact with their environments and how to reduce vulnerabilities. This knowledge helps guide investments in

infrastructure, communication, and governance that strengthen collective capacity to withstand disruptions. This thesis contributes to the field by integrating mathematical theory, computational simulation, and deep learning methods into a unified framework for studying tipping points. The theoretical component establishes connections between bifurcation theory, stochastic processes, and resilience indicators. The simulation component generates synthetic data from canonical models and coupled socio-climate systems, providing a controlled environment in which to test hypotheses. The deep learning component leverages these datasets to train algorithms capable of detecting and classifying critical transitions in complex, noisy settings. By bridging these domains, the thesis advances beyond traditional approaches that rely solely on analytical methods or empirical case studies. The hybrid strategy demonstrates how modern computational tools can extend classical theory, capturing richer patterns of instability and providing more robust predictions. Importantly, it also shows how deep learning models trained on simplified simulations can generalise to real-world data, creating new opportunities for application. The overarching contribution is to provide both methodological innovation and practical insight. By linking theory to data-driven techniques, the thesis offers a pathway toward more accurate, interpretable, and actionable early warning systems. These advances not only deepen scientific understanding of complex dynamics but also enhance society’s ability to anticipate and respond to critical transitions in climate, ecosystems, and human systems.

1.7.1 Aims and Objectives

The overarching aim of this thesis is to develop, test, and apply new methods for understanding and anticipating critical transitions in complex dynamical systems, with a particular focus on socio-climate interactions. Traditional approaches to tipping points have often relied either on highly simplified analytical models or on empirical statistical indicators that may fail under realistic noise conditions and limited data availability. This thesis seeks to bridge that gap by combining mathematical theory, computational simulations, and modern deep learning methods into a unified framework.

More specifically, the objectives of this work are as follows:

- To formalise and analyse early warning signals of tipping points using normal form theory, stochastic processes, and bifurcation analysis, thereby providing a rigorous theoretical basis for prediction.
- To design deep learning architectures capable of classifying different types of bifurcations and detecting flickering phenomena that precede collapse, under both white and coloured noise.

- To integrate statistical early warning signals with machine learning predictions, improving robustness when working with noisy, short, or irregular datasets.
- To develop coupled socio–climate models that capture how human mitigation behaviours interact with climate tipping elements, including feedback loops that may accelerate or delay collapse.
- To compare binary versus spectrum opinion models in socio–climate dynamics, clarifying the trade-off between simplicity and descriptive richness in representing collective human behaviour.
- To provide actionable insights that can inform risk management, climate policy, and resilience strategies by demonstrating how new computational tools can be linked to pressing societal challenges.

1.7.2 Thesis Outline

This thesis is structured to gradually build from theoretical foundations, through methodological development, to applied socio–climate case studies. Each chapter addresses a distinct research question, but they are unified by the common goal of advancing early detection of critical transitions.

- **Chapter 2** develops a deep learning framework for detecting bifurcations in dynamical systems with coloured noise. The chapter explains model construction, training dataset generation, and performance benchmarks against classical early warning signals.
- **Chapter 3** extends the deep learning approach to detect flickering, a noise-induced precursor of collapse. This chapter demonstrates how data-driven tools can distinguish genuine early warnings from confounding noise, with applications to both synthetic and empirical datasets.
- **Chapter 4** introduces coupled socio–climate models in which human behaviour and climate tipping elements interact. It shows how social dynamics can either delay or accelerate climate tipping points, thereby highlighting the critical role of behavioural feedbacks.
- **Chapter 5** investigates whether simplified binary social models are sufficient to capture socio–climate complexity, or whether spectrum models provide fundamentally different insights. This comparative analysis clarifies when simplicity is justified and when richer representations are necessary.

- **Conclusion** synthesizes the key findings across all chapters, highlighting how deep learning, dynamical systems analysis, and socio–climate modeling jointly advance our understanding of early warning signals and structural change. It also outlines limitations and promising directions for future research.
- **Appendices** provide supplementary material, including detailed derivations, sensitivity analyses, extended model descriptions, and data/code availability statements. These resources are intended to support replication and further research.

Together, these chapters aim to demonstrate that by combining classical dynamical systems theory with modern deep learning methods, it is possible to create a more robust and versatile toolkit for anticipating critical transitions in both natural and socio–climate systems.

Chapter 2

Deep Learning for Bifurcation Detection: Extending Early Warning Signals to Dynamical Systems with Coloured Noise¹

¹This chapter has been peer reviewed and published as: Babazadeh, Y.; Dylewsky, D.; Anand, M.; Bauch, C.T. *Mathematics* **2025**, **13**(17), Section: Applied Mathematics. Publication date: 29 August 2025.

2.1 Abstract

Deep learning models have demonstrated remarkable success in recognising tipping points and providing early warning signals. However, there has been limited exploration of their application to dynamical systems governed by coloured noise, which characterizes many real-world systems. In this study, we show that it is possible to leverage the normal forms of three primary types of bifurcations (Fold, Transcritical, and Hopf) to construct a training set that enables deep learning architectures to perform effectively. Furthermore, we showed that this approach could accommodate coloured noise by replacing white noise with red noise during the training process. To evaluate the classifier trained on red noise compared to one trained on white noise, we tested their performance on mathematical models using Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) scores. Our findings reveal that the deep learning architecture can be effectively trained on colored noise inputs, as evidenced by high validation accuracy and minimal sensitivity to redness (ranging from 0.83 to 0.85). However, classifiers trained on white noise also demonstrate impressive performance in identifying tipping points in colored time series. This is further supported by high AUC scores (ranging from 0.9 to 1) for both classifiers across different colored stochastic time series..

2.2 Introduction

Dynamical systems are employed to describe various phenomena across disciplines, from studying the climate system [30, 122, 10] to epidemiological studies [208, 115, 12], rumor propagation mechanism [178], and coupled behavior-disease dynamics [80]. While dynamical systems exhibit numerous intriguing aspects, one particularly captivating focus has been the possibility of bifurcation and tipping points [175, 65, 80, 68]. In natural systems, a local bifurcation can occur as external conditions gradually change, bringing the system near an equilibrium state and potentially triggering a qualitative shift in behaviour when a tipping point is reached [136, 155].

Bifurcation theory explores how small changes in system parameters can lead to sudden qualitative transitions in the behavior of dynamical systems. [40]. One of its most significant theories, the Centre Manifold Theorem, along with the Normal Form Theorem, simplifies nonlinear dynamical systems by transforming them into their simplest equivalent form, termed the "normal form". Through coordinate transformations, it eliminates unnecessary nonlinear terms while preserving the system's essential behaviour, making it easier to examine features like bifurcations [102, 198, 29].

While bifurcation theory encompasses a broad set of rigorous theorems characterizing

qualitative changes in dynamical systems, this work specifically focuses on three types of local bifurcations, as they represent distinct behaviours in dynamical systems: the abrupt jump from one equilibrium state to another (Fold bifurcation) [114], smoothly shifting stability (Transcritical bifurcation) [151], and the transition to oscillatory behaviour (Hopf bifurcation) [142]. For these bifurcation types, the dominant eigenvalue of the derivative matrix crosses the imaginary axis at the bifurcation point. Leading up to the bifurcation, the system can exhibit critical slowing down, corresponding to prolonged recovery from perturbations and reduced resilience to disturbances, as measured by increased auto-correlation and variance [200]. Numerous generic indicators of local bifurcations have been proposed based on these consequences [32, 42, 45, 57]. For instance, the variance and lag-1 auto-correlation of a time series prior to bifurcation shows trends that can serve as early warning signals (EWS) [41]. While such indicators have been instrumental in identifying and predicting tipping points and abrupt shifts, they also possess limitations [44], such as an inability to specify the type of impending bifurcation. With recent advances in artificial intelligence (AI) and the development of deep learning (DL) models, it was anticipated that these models could eventually provide EWS with greater accuracy and fewer weaknesses than generic indicators.

The first study applying AI to the detection of tipping points in dynamical systems is by Bury et al.[34], who showed that convolutional and recurrent neural networks could identify EWS in time series drawn from a random library of differential equations. Their approach demonstrated that DL architectures could outperform traditional statistical indicators. Deb et al.[46] introduced EWSNet, a hybrid Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) network trained on simulated time series from ecological, biological, and climate models. EWSNet showed high accuracy in classifying catastrophic and non-catastrophic transitions and proved robust to noise, short time series, and irregular sampling. These previous works were extended in a follow-up study by Bury et al.[33] to discrete-time systems, further validating the generalizability of DL models to diverse bifurcation scenarios. Dylewsky et al.[59] demonstrated that universal EWS could be extracted from DL models trained on climate system simulations, supporting the broader applicability of these methods. Dylewsky et al.[57] advanced this work by embedding bifurcations in high-dimensional systems, showing that DL models remain effective even when the underlying signals are distributed across many variables. Most recently, Huang et al.[94] addressed the challenge of rate-induced tipping (R-tipping), where classical indicators like critical slowing down fail. Their interpretable DL framework predicted transition probabilities in noisy, time-varying systems and extracted higher-order fingerprints of instability. In parallel, Dylewsky et al. [58] tackled spatially patterned phase transitions using neural architectures tailored for spatiotemporal systems. Together, these studies illustrate a growing body of work that supports the use of deep learning, with CNN–LSTM

architectures being one of the main and most effective models for robust and generalizable detection of critical transitions in noisy, nonlinear dynamical systems.

In contrast to previous approaches that generate extensive libraries of stochastic differential equations and discard those that do not exhibit bifurcations [33], we propose a theoretically grounded method by constructing the training dataset directly from the normal forms of canonical bifurcations. This significantly reduces computational overhead while ensuring that the training data captures the essential dynamical features associated with each bifurcation type. The first objective of this study is to demonstrate that such a targeted construction is not only feasible but also more efficient than relying on random dynamical systems. Furthermore, we extend the DL-based early warning framework of [34] to accommodate coloured noise, a feature often overlooked despite its prevalence in empirical systems. Many ecological and climate systems that undergo critical transitions are influenced by auto-correlated, coloured noise [158, 131], prompting recent interest in its ecological impacts [76, 75, 192, 73, 132, 202, 174, 163, 173, 204]. By generating training data with varying levels of redness, we evaluate the robustness of our CNN-LSTM classifier under both white and coloured noise, enhancing the applicability of EWS frameworks to real-world, noisy, nonlinear systems.

2.3 Methods

2.3.1 Constructing training data set

The training dataset was created by stochastically simulating the normal forms of three canonical local bifurcations, fold, transcritical, and Hopf, each with an additive stochastic term. An additional “Null” category was generated by simulating stochastic noise alone, serving as a baseline non-bifurcating system. In each case, the system was initialized at an equilibrium point far from the bifurcation and driven closer to the bifurcation point through linear parameter variation. The resulting time series were classified into four categories (fold, transcritical, Hopf, null) and used to train the CNN-LSTM model.

Normal Form Equations

We used the following stochastic differential equations to construct the dataset:

Fold bifurcation:

$$\frac{dx}{dt} = \mu + x^2 + \sigma \cdot \epsilon_t, \quad (2.1)$$

with the bifurcation occurring at $\mu = 0$.

Transcritical bifurcation:

$$\frac{dx}{dt} = rx + x^2 + \sigma \cdot \epsilon_t, \quad (2.2)$$

where the bifurcation occurs at $r = 0$.

Hopf bifurcation:

$$\frac{dz}{dt} = (\mu + i\omega)z - |z|^2 z + \sigma \cdot \epsilon_t, \quad (2.3)$$

where the bifurcation occurs at $\mu = 0$.

Null case:

$$\frac{dx}{dt} = \sigma \cdot \epsilon_t. \quad (2.4)$$

Here, σ denotes the noise intensity, and ϵ_t is a stochastic term that can be either white noise ($\epsilon_t \sim N(0, 1)$) or red noise generated by the AR(1) process:

$$X_{t+1} = (\text{redness} \cdot X_t) + \sqrt{1 - \text{redness}^2} \cdot \epsilon_t, \quad (2.5)$$

with $X_0 \sim N(0, 1)$. The parameter *redness* controls the autocorrelation: *redness* = 0 corresponds to white noise, while higher values yield stronger temporal correlations. In the limiting case, when *redness* = 0, the process produces a purely uncorrelated time series of white noise samples, whereas when *redness* = 1, the process becomes fully persistent, resulting in a constant time series equal to its initial value X_0 , which is drawn from the normal distribution $N(0, 1)$.

Simulation Procedure

For each bifurcation type, we simulated 5000 time series, each consisting of 1000 time steps. The control parameter (μ or r) was linearly varied from an initial value far from the bifurcation point (e.g., $\mu \in [-3, -2]$ for fold bifurcation, with $x_0 = \sqrt{-\mu}$) toward the critical point. To ensure diversity and robustness in the training dataset, the initial values of the bifurcation parameter were randomly selected from their respective ranges, and the noise amplitude σ was sampled uniformly from the interval $[0.01, 0.1]$. This stochastic sampling of both system and noise parameters enabled the generation of a heterogeneous set of trajectories, reflecting variability in dynamical regimes and noise intensities across simulations. The simulation was halted before reaching the bifurcation threshold to ensure that the DL model was exposed only to the pre-bifurcation regime. The Euler–Maruyama method was used for numerical integration with a fixed step size of $\Delta t = 0.01$. This procedure was repeated for six noise colours, corresponding to redness values of 0.0, 0.1, 0.2, 0.3, 0.4, and 0.5.

Pre-processing and Model Training

Before being fed into the CNN-LSTM network, the simulated time series were pre-processed using a Gaussian kernel to smooth stochastic fluctuations while preserving dynamical structure. The kernel width was set to $0.1 \times L$, where $L = 1000$ is the length of each time series. Other pre-processing techniques (moving average, first differencing, Z-score normalization) were tested, but the Gaussian kernel yielded the best classification performance.

In total, 20,000 time series were generated. Ninety percent were used for training, and the remaining ten percent for testing and validation. The model’s performance across different redness values is summarized in Table 2.1.

We used these mathematical metrics for the evaluation of the deep learning model’s classification performance across different noise conditions. The redness parameter quantifies the degree of temporal autocorrelation in the input noise and ranges from 0 (white noise) to 0.5 (correlated red noise). Higher redness values indicate increased memory and persistence in the noise process, as defined by the AR(1) formulation in Equation 2.5.

The **validation accuracy** and **accuracy score** measure the overall proportion of correctly predicted samples, defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP , TN , FP , and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively. This metric captures general correctness across all classes but can be insensitive to class imbalance.

To capture both the reliability and completeness of positive predictions, we computed the **precision** and **recall** metrics. Precision measures the fraction of correctly predicted positive samples among all positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP},$$

while recall (also known as sensitivity) quantifies the fraction of true positive cases correctly identified by the model:

$$\text{Recall} = \frac{TP}{TP + FN}.$$

The **F1 score** provides a balanced measure that combines precision and recall as their harmonic mean:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

A higher F1 score indicates that the model maintains both high precision and high recall, avoiding bias toward over-prediction or under-detection of any class.

Together, these complementary metrics provide a comprehensive quantitative assessment of model performance, ensuring that both overall accuracy and class-specific sensitivity are robust across varying levels of noise persistence.

Table 2.1: Model Performance Metrics Across Redness Levels

Redness	Validation Accuracy	Accuracy Score	F1 Score	Precision	Recall
0.0	0.852	0.826	0.825	0.829	0.827
0.1	0.857	0.831	0.829	0.832	0.832
0.2	0.857	0.833	0.834	0.836	0.837
0.3	0.853	0.834	0.836	0.848	0.836
0.4	0.831	0.841	0.838	0.847	0.840
0.5	0.832	0.864	0.864	0.866	0.864

Training process information

The **confusion matrix** is a fundamental tool for evaluating classification performance in supervised learning. It provides a detailed breakdown of how many samples from each true class were correctly or incorrectly classified by the model. For a multi-class problem with n categories, the confusion matrix is an $n \times n$ table where each element C_{ij} represents the number of observations belonging to the true class i that were predicted as class j . The diagonal elements C_{ii} correspond to correctly classified samples (true positives for each class), while the off-diagonal elements represent misclassifications. In the context of our study, this matrix allows us to identify which bifurcation types (fold, transcritical, Hopf, or null) are more prone to being confused with one another across different noise conditions.

From the confusion matrix, we derive several **classification metrics**, including precision, recall, F1 score, and overall accuracy, that quantify the model’s predictive performance. These metrics summarize the trade-offs between false positives and false negatives, providing complementary perspectives on model robustness and class separability. In particular, precision measures the reliability of positive predictions, recall quantifies the model’s sensitivity to true cases, and the F1 score balances both aspects. Together, the confusion and classification matrices (Tables 2.2 and 2.3) provide a comprehensive diagnostic of the CNN–LSTM model’s ability to distinguish between pre-bifurcation dynamics associated with different normal forms under varying noise persistence.

Table 2.2: Confusion matrices for CNN-LSTM classifier performance across redness levels. Each matrix shows predicted (columns) versus true (rows) class counts.

Redness = 0	Redness = 0.1
$\begin{bmatrix} 249 & 0 & 0 & 3 \\ 0 & 205 & 26 & 18 \\ 0 & 28 & 191 & 14 \\ 2 & 49 & 34 & 181 \end{bmatrix}$	$\begin{bmatrix} 252 & 1 & 0 & 1 \\ 0 & 207 & 12 & 18 \\ 0 & 34 & 184 & 28 \\ 4 & 39 & 32 & 188 \end{bmatrix}$
Redness = 0.2	Redness = 0.3
$\begin{bmatrix} 238 & 1 & 0 & 1 \\ 0 & 213 & 15 & 13 \\ 0 & 33 & 183 & 29 \\ 6 & 40 & 29 & 199 \end{bmatrix}$	$\begin{bmatrix} 228 & 1 & 0 & 1 \\ 0 & 245 & 11 & 15 \\ 0 & 47 & 176 & 24 \\ 5 & 46 & 16 & 185 \end{bmatrix}$
Redness = 0.4	Redness = 0.5
$\begin{bmatrix} 246 & 1 & 1 & 1 \\ 0 & 234 & 12 & 6 \\ 0 & 39 & 192 & 23 \\ 7 & 41 & 28 & 169 \end{bmatrix}$	$\begin{bmatrix} 246 & 0 & 0 & 1 \\ 0 & 206 & 15 & 31 \\ 0 & 18 & 197 & 32 \\ 5 & 10 & 24 & 215 \end{bmatrix}$

Robustness of training process

To evaluate the stability and reproducibility of our model, we trained the same CNN-LSTM architecture 10 times using the same training dataset containing white noise and identical hyperparameter settings. Table 2.4 reports the mean and standard deviation of the key performance metrics, accuracy, F1 score, precision, and recall, across these repeated runs. The results show strong and consistent performance, with mean values centered around 0.82 and standard deviations ranging from 0.009 to 0.010. This low variability across identical training conditions demonstrates that the model’s performance is robust to random initialization and other sources of stochasticity inherent in the training process, reinforcing the reliability of our deep learning framework.

Table 2.3: Classification metrics (precision, recall, and F1-score) for the CNN-LSTM classifier under different redness values.

Class	Redness = 0			Redness = 0.1		
	Prec.	Rec.	F1	Prec.	Rec.	F1
Fold	0.992	0.988	0.990	0.984	0.992	0.988
Transcritical	0.727	0.823	0.772	0.737	0.873	0.799
Hopf	0.761	0.820	0.789	0.807	0.748	0.776
Null	0.838	0.680	0.751	0.800	0.715	0.755
Class	Redness = 0.2			Redness = 0.3		
	Prec.	Rec.	F1	Prec.	Rec.	F1
Fold	0.975	0.992	0.983	0.979	0.991	0.985
Transcritical	0.742	0.884	0.807	0.723	0.904	0.803
Hopf	0.806	0.747	0.775	0.867	0.713	0.782
Null	0.822	0.726	0.771	0.822	0.734	0.776
Class	Redness = 0.4			Redness = 0.5		
	Prec.	Rec.	F1	Prec.	Rec.	F1
Fold	0.972	0.988	0.980	0.980	0.996	0.988
Transcritical	0.742	0.929	0.825	0.880	0.817	0.848
Hopf	0.824	0.756	0.789	0.835	0.798	0.816
Null	0.849	0.690	0.761	0.771	0.846	0.807

Data Requirements and Sequence Length Sensitivity

In the left panel of Figure 2.1, we evaluate how the amount of training data affects model performance. Starting with a dataset of 20,000 time series, we incrementally increase the fraction used for training while keeping the test set fixed, and assess the model’s performance using the F1 score. We observe a significant improvement in performance as the training fraction increases from 0.1 to 0.7, with the F1 score rising sharply in the early stages. Beyond a training fraction of 0.7, the performance plateaus, indicating that the model reaches a saturation point in learning. This behaviour suggests that the training process is robust and that a subset of the full dataset (around 70 percent) is sufficient to achieve near-optimal performance, highlighting the data efficiency of our DL framework.

In the right panel of Figure 2.1, we examine the impact of time series length on model precision. Using the full set of 20,000 time series, we simulate the same experimental

Table 2.4: Model performance metrics with mean and standard deviation over multiple runs.

Metric	Mean Score	Standard Deviation
Accuracy	0.819	± 0.009
F ₁ Score	0.819	± 0.010
Precision	0.826	± 0.010
Recall	0.821	± 0.009

setting across various sequence lengths to test the model’s sensitivity to temporal resolution. The results show that for relatively short time series, model precision is notably lower, likely due to insufficient temporal information to capture the dynamics preceding a bifurcation. However, as the sequence length increases, performance improves steadily, eventually plateauing beyond 800 time steps. This indicates a minimum required length for capturing the relevant precursors to tipping points. While longer sequences beyond this threshold do not significantly improve performance, they may be advantageous when training classifiers for use with higher-resolution simulated or empirical datasets.

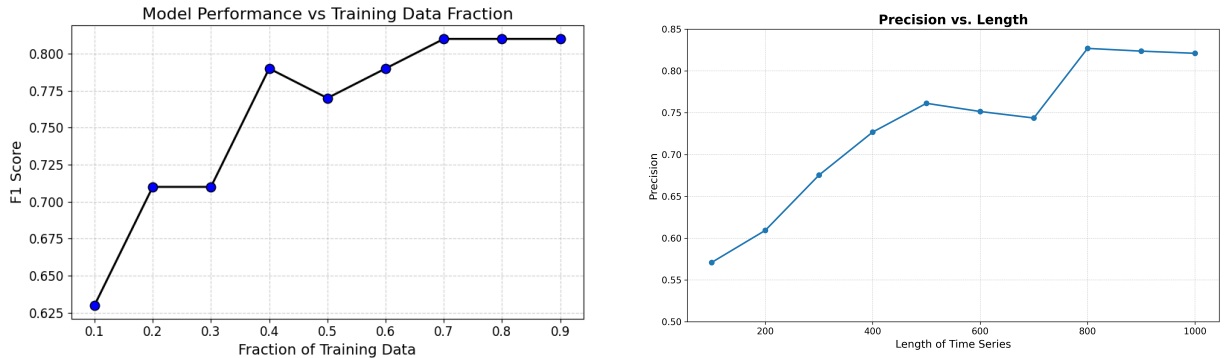


Figure 2.1: (Left) Model performance as a function of training data fraction. (Right) Precision as a function of time series length.

2.3.2 Model Selection Rationale

To construct the deep learning model, we employed a hybrid architecture combining CNN with LSTM network [177, 3, 88]. Theoretically, the CNN layers serve to extract

local patterns and short-term features from the time series by applying convolutional filters, while the LSTM layers are designed to capture long-range temporal dependencies through their recurrent structure. This layered combination is particularly well-suited for our task, as it enables both local pattern detection and memory retention, key components in identifying EWS indicative of critical transitions.

The CNN-LSTM architecture has been widely adopted in the broader literature on time series forecasting and classification. For instance, [135] demonstrated that CNN-LSTM achieved the highest forecasting accuracy across multiple benchmarks, outperforming standalone models such as Multilayer Perceptron (MLP), CNN, Recurrent Neural Network (RNN), and LSTM. Similarly, [137] evaluated CNN, LSTM, and CNN-LSTM on ice-jam event classification and reported superior F1-scores for the CNN-LSTM model. In the context of missing data imputation, [96] showed that CNN-LSTM achieved better predictive accuracy than either CNN or LSTM alone when forecasting electricity consumption. More specifically, architectures similar to ours have been increasingly applied to problems involving bifurcation detection and phase transitions in dynamical systems. Recent studies such as [58, 57, 33, 59, 34] have demonstrated that CNN-LSTM models are effective at detecting EWS in stochastic systems and outperforming simpler baselines.

To justify our choice of DL model, we conducted an ablation study comparing the performance of LSTM-only and CNN-LSTM architectures. Both models were trained and tested on time series generated with white noise, using identical hyperparameter settings to ensure a fair comparison. To ensure a fair evaluation, we tested both the LSTM-only and CNN-LSTM models on time series generated using the same process as the training dataset. The results, presented in Table 2.5, show that while the LSTM-only model achieves reasonably strong performance, the CNN-LSTM consistently outperforms it in terms of accuracy, precision, recall, and F1 score. This supports our choice of architecture and aligns with findings from previous work, such as [34], which also highlighted the superiority of CNN-LSTM over LSTM-only models in similar dynamical settings. Both the CNN-LSTM and LSTM-only models were trained on the same dataset from normal forms with white noise, using identical settings and preprocessing. The LSTM-only model shares the recurrent layers of the CNN-LSTM but omits the convolutional block. Hyperparameters were guided by [34] and further tuned through experiments to optimize performance. More details on the DL architecture are provided in Appendix A1.

Table 2.5: Comparison of overall performance metrics between CNN-LSTM and LSTM-only models.

Metric	CNN-LSTM	LSTM-only
Accuracy Score	0.823	0.773
F1 Score	0.822	0.769
Precision	0.834	0.778
Recall	0.823	0.773

2.4 Results

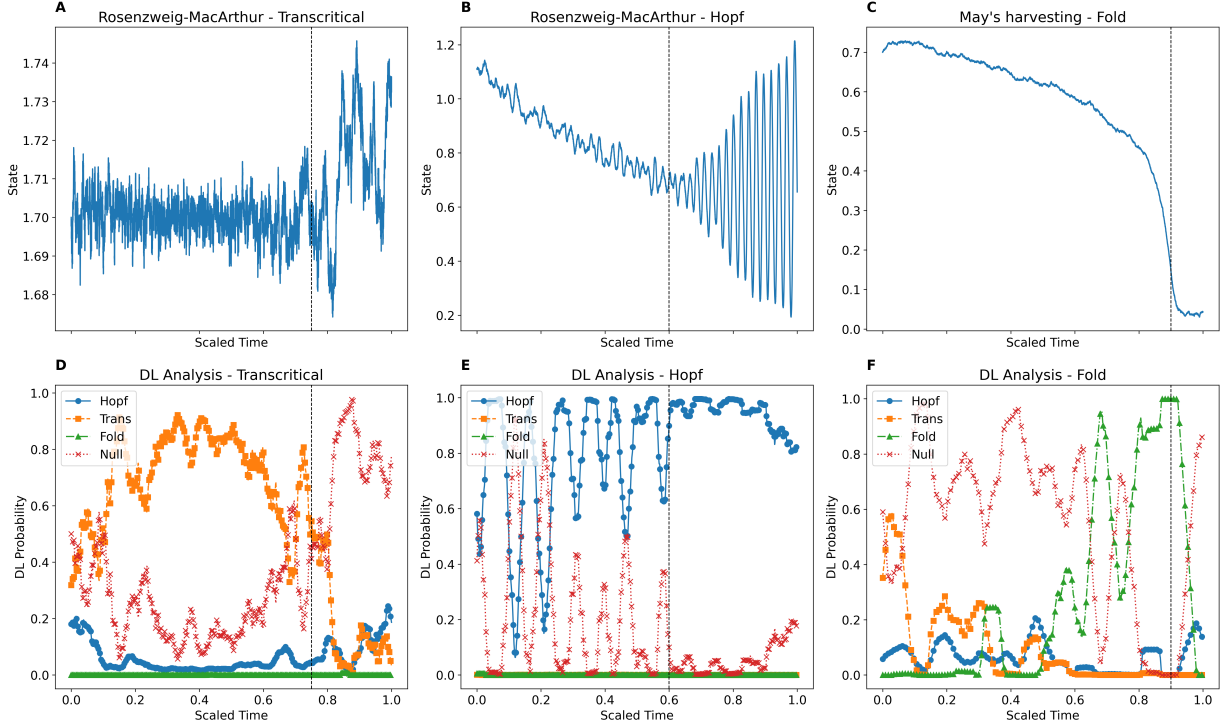


Figure 2.2: **DL model providing EWS prior to bifurcation points for Rosenzweig-MacArthur and May's harvesting model.** (A–C) Synthetic time series from three canonical bifurcation models with additive coloured noise: (A) Rosenzweig-MacArthur predator-prey system near a transcritical bifurcation; (B) Rosenzweig-MacArthur with Hopf bifurcation; (C) May's harvesting model near a fold bifurcation. (D–F) DL model predictions for the systems in (A–C), showing temporal evolution of bifurcation class probabilities. Probabilities are represented with distinguishable markers and line styles for print clarity: triangle (fold), square (transcritical), circle (Hopf), and x (null). The vertical dashed line indicates the approximate location of the tipping point.

Evaluating the DL Model on Mathematical Systems

In this section, we aim to demonstrate that the trained deep learning (DL) model can provide early warning signals (EWS) for bifurcations in well-known dynamical systems. For this purpose, we employed six mathematical models commonly used to study bifurcation

phenomena: the Neural Activation model and May’s Harvesting model for fold bifurcations; the Rosenzweig–MacArthur and SIS models for transcritical bifurcations; and the Rosenzweig–MacArthur and Van der Pol models for Hopf bifurcations.

Each model was simulated with an additive stochastic term, starting from its stable equilibrium and linearly varying the corresponding bifurcation parameter toward the critical point throughout the simulation. Because tipping points can occur earlier than the theoretical bifurcation threshold, we employed the binary segmentation algorithm to identify the empirical location of the tipping point [106]. The DL model then assigned four probabilities (Fold, Transcritical, Hopf, and Null) to each time step in the simulated series. The Null category represents stochastic time series that do not contain a tipping point or are far from the bifurcation region. Assigning a high probability to the correct bifurcation type before the actual tipping point is considered an effective early warning signal in this study.

Mathematical Models

May’s Harvesting Model (Fold Bifurcation):

$$\frac{dx}{dt} = rx(1 - x) - h \frac{x^2}{x^2 + s^2},$$

where x represents population size, r is the intrinsic growth rate, and h is the harvesting rate. The term $\frac{x^2}{x^2 + s^2}$ introduces a saturating effect in the harvesting process, with s controlling the steepness of saturation. A fold bifurcation occurs as h is varied.

Neural Activation Model (Fold Bifurcation):

$$\frac{dx}{dt} = -x + \frac{S}{1 + e^{-\beta(x-\theta)}},$$

where x denotes the activation level, S is the external stimulus, β is the steepness of the sigmoid response, and θ is the threshold. A fold bifurcation arises as S is increased.

Rosenzweig–MacArthur Consumer–Resource Model (Transcritical Bifurcation):

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{K}\right) - \frac{axy}{b + x}, \quad \frac{dy}{dt} = c \frac{axy}{b + x} - dy,$$

where x and y denote the resource and consumer populations, respectively. Here, r is the intrinsic resource growth rate, K is the carrying capacity, a is the maximum predation rate, b is the half-saturation constant, c is the conversion efficiency, and d is the consumer death rate. A transcritical bifurcation occurs as d is varied.

Rosenzweig–MacArthur Consumer–Resource Model (Hopf Bifurcation): The same system of equations applies here, but a Hopf bifurcation occurs when d crosses a critical threshold, producing sustained oscillations in both populations.

Van der Pol Oscillator (Hopf Bifurcation):

$$\frac{dx}{dt} = y, \quad \frac{dy}{dt} = \mu(1 - x^2)y - x,$$

where x and y are the state variables and μ is the bifurcation parameter. A Hopf bifurcation occurs at $\mu = 0$, where the system transitions from a stable equilibrium to periodic oscillations.

SIS Epidemic Model (Transcritical Bifurcation):

$$\frac{dS}{dt} = -\beta SI + \gamma I, \quad \frac{dI}{dt} = \beta SI - \gamma I,$$

where S and I represent the susceptible and infected populations, respectively, β is the infection rate, and γ is the recovery rate. A transcritical bifurcation occurs at the critical threshold $R_0 = \frac{\beta}{\gamma} = 1$, where R_0 denotes the basic reproduction number.

Simulation and DL Analysis

For each model, we integrated the stochastic differential equations using the Euler–Maruyama method with $\Delta t = 0.01$. The DL model analyzed each simulated time series using a sliding window of 100 time steps, assigning probabilities for each segment to belong to one of the four classes. To smooth short-term fluctuations in the output probabilities, we applied a moving average with a window size of 100 as a post-processing step.

Figures 2.2 and 2.3 illustrate representative results of the DL analysis for the different mathematical models. The estimated tipping points, detected via binary segmentation, are indicated by dashed lines. As shown, the DL model not only provides reliable EWS but also accurately identifies the correct type of bifurcation. At the beginning of each simulation, when the system is far from the critical threshold, the DL model assigns the highest probability to the Null class. As the system approaches the tipping point, the probability of the correct bifurcation type rises sharply and then returns to Null after the transition has passed. This consistent behavior across all six models confirms the DL model’s capacity to detect and classify bifurcation-induced transitions. More details on the validation on bifurcating and null dynamical systems are provided in Appendix A2.

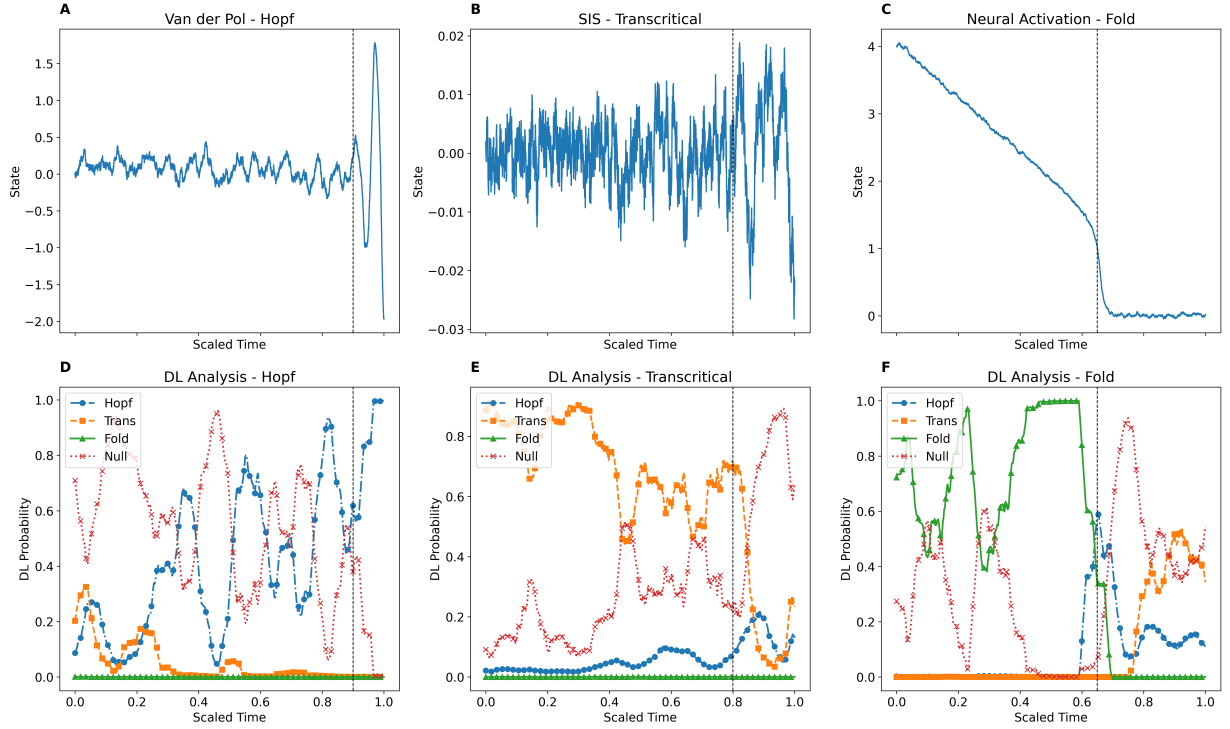


Figure 2.3: DL model providing EWS prior to bifurcation points for Van der Pol, SIS, and Neural activation model. (A–C) Synthetic time series from three canonical bifurcation models with additive white noise: (A) Van der Pol oscillator approaching a Hopf bifurcation; (B) SIS model undergoing a transcritical bifurcation; (C) Neural activation model with a fold bifurcation. The vertical dashed lines indicate the location of the bifurcation. (D–F) Output of the DL classifier showing predicted bifurcation probabilities over time for the corresponding systems in (A–C). Distinct line styles and markers denote bifurcation types: triangle (fold), square (transcritical), circle (Hopf), and x (null).

Application to Empirical Data

In this study, we tested the performance of the trained deep learning (DL) classifier on two distinct sources of empirical data that exhibit critical transitions. These datasets represent fundamentally different physical systems, one geochemical and the other thermoacoustic, allowing us to assess the generality and robustness of the model in detecting early warning signals (EWS).

Sedimentary Archives (Mediterranean Sea). This dataset provides high-resolution reconstructions of oxygen dynamics from sediment cores in the eastern Mediterranean Sea [84]. It captures transitions between oxic and anoxic states, where early warning signals have been previously identified [85]. The data consist of molybdenum (Mo) and uranium (U) concentrations, which act as proxies for anoxic and suboxic conditions, respectively. Across three sediment cores, eight distinct anoxic events were identified, yielding a total of 26 time series with temporal resolutions ranging from 10 to 50 years, depending on the core.

Thermoacoustic Instability (Rijke Tube Experiments). This dataset investigates transitions to self-sustained oscillations driven by thermoacoustic instability [161]. Experiments were conducted in a horizontal Rijke tube, where a heated wire mesh generated controlled instabilities through voltage variation, inducing subcritical Hopf bifurcations. The data include 19 forced trajectories at different voltage ramp rates and 10 steady-state trajectories at fixed voltages. Each time series, originally sampled at 4–10 kHz, was downsampled to 2 kHz, retaining approximately 1,500 data points prior to each transition.

DL Model Analysis on Empirical Data

To ensure consistent temporal resolution and facilitate DL-based analysis, all empirical time series were rescaled to a common length using linear interpolation. This method estimates intermediate data points by connecting consecutive observations with straight lines, thereby preserving the overall trend and continuity of the original signal without introducing artificial oscillations [167]. Each time series was then analyzed using the trained CNN-LSTM model, which assigns probabilities corresponding to the four possible classes (Fold, Transcritical, Hopf, and Null) for each time step. The results were smoothed using a moving average to reduce short-term fluctuations.

Figure 2.4 illustrates representative DL outputs for four example time series (two from each dataset). The estimated tipping point, detected via binary segmentation, is indicated by a dashed line in each subplot. As shown, the DL model successfully provides early warning signals prior to the observed transitions, with a clear shift in classification probabilities from Null to one of the bifurcation types.

Interestingly, for the sedimentary archive data, the DL model frequently identifies the transition as Hopf-type, whereas empirical and geochemical evidence suggests that these events more closely resemble fold-type bifurcations. This discrepancy may reflect the model’s sensitivity to oscillatory components in the noisy pre-transition data. Nevertheless, the DL classifier consistently detects approaching transitions across both datasets,

demonstrating robust generalization to real-world systems characterized by different noise structures, time scales, and underlying mechanisms.

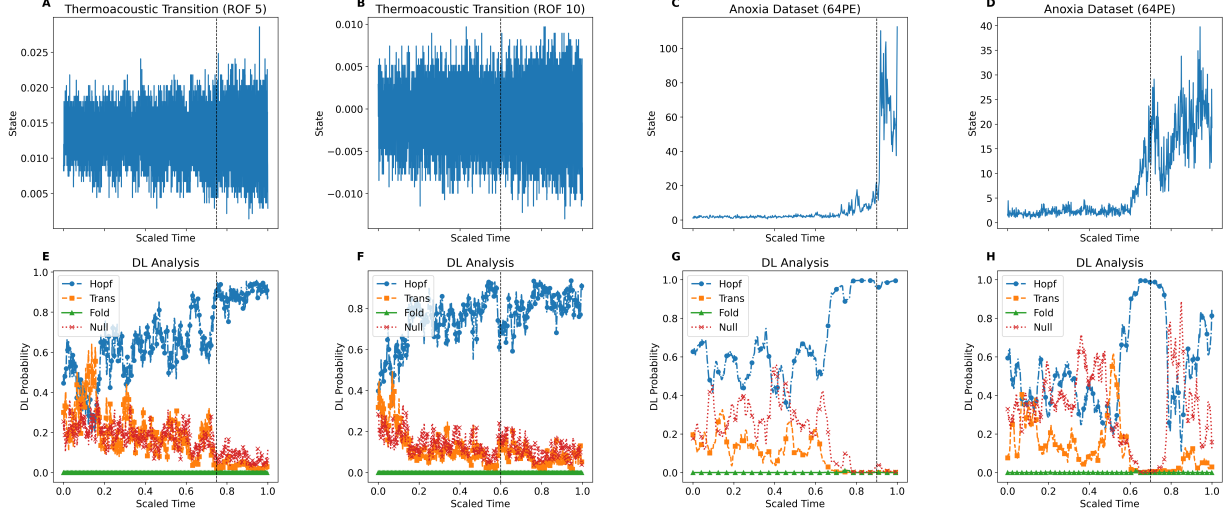


Figure 2.4: DL model providing EWS before bifurcation points. (A–D) Real-world time series examples: (A–B) Thermoacoustic instability signals from combustor experiments at ROF 5 and ROF 10; (C–D) Geological anoxia indicators from sediment cores at two locations. (E–H) Corresponding DL-based bifurcation predictions for (A–D), with line styles and markers indicating the inferred probability of each class. This illustrates how the model generalizes from synthetic to empirical data. Distinct line styles and markers denote bifurcation types: triangle (fold), square (transcritical), circle (Hopf), and x (null). The vertical dashed lines are the approximate location of the tipping point.

Classifier Performance Across Models and Redness Values

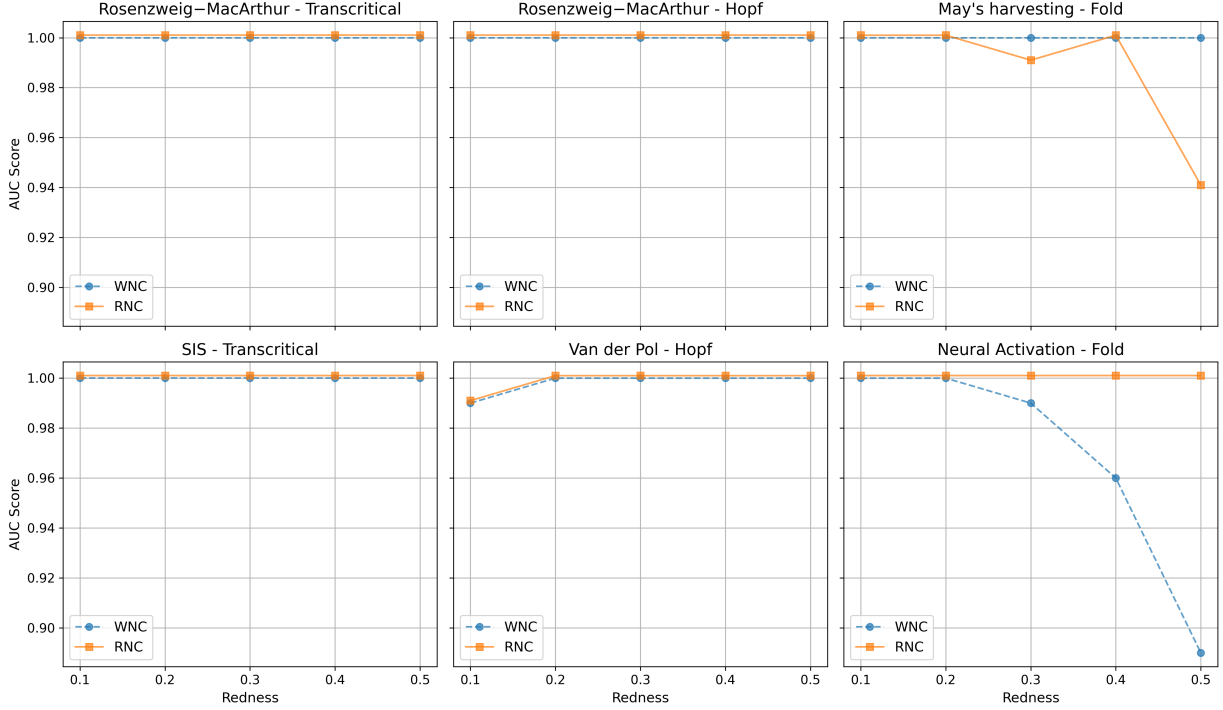


Figure 2.5: **AUC scores of White Noise and Red Noise classifiers across six mathematical models** Classifier performance across six dynamical systems under varying levels of noise correlation (redness). Each subplot displays the AUC score as a function of redness for white noise conditions (WNC, blue circles) and red noise conditions (RNC, orange squares). **Top row:** Rosenzweig-MacArthur model with transcritical and Hopf bifurcations, and May's harvesting model with a fold bifurcation. **Bottom row:** SIS model (transcritical), Van der Pol oscillator (Hopf), and neural activation model (fold). The results highlight that while most systems retain high classification accuracy across redness levels, performance in certain fold bifurcations, especially under WNC in the neural activation model and under RNC in the May model, shows degradation as temporal correlation increases.

Robustness to Coloured Noise and ROC-AUC Evaluation

In this section we aim to illustrate that the DL model demonstrates consistent success with stochastic time series that include coloured noise. To illustrate this, we repeated the

same training process using the same DL architecture (CNN–LSTM), but this time incorporated red noise as defined by the AR(1) process in Equation 2.5 as an additive stochastic term while simulating the deterministic normal forms of different types of bifurcation. As a result, we generated five additional DL classifiers, each trained on normal forms combined with red noise of differing autocorrelation strengths, corresponding to redness values of 0.1, 0.2, 0.3, 0.4, and 0.5. A larger redness value signifies a stronger dominance of low-frequency components, greater persistence, and smoother long-term variability in the time series. The baseline DL model used in previous sections corresponds to the white-noise case (*redness* = 0), which can be viewed as a special instance of this family of models.

To assess and compare the discriminative performance of the classifiers, we computed Receiver Operating Characteristic (ROC) curves across a range of simulated dynamical systems. Each system was configured to either gradually approach a bifurcation point (signal present) or remain in a stable regime with fixed parameters (signal absent). This binary labeling allowed quantitative evaluation of the classifier’s ability to distinguish between pre-bifurcation and null dynamics. Each time series was then classified using two CNN–LSTM models: one trained on white noise ($r = 0$) and one trained on red noise generated using an AR(1) process with varying autocorrelation levels ($r \in [0.1, 0.5]$). The model outputs softmax probabilities for the four classes: Fold, Transcritical, Hopf, and Null.

For ROC construction, we treated the mean probability of the correct bifurcation class as the classifier score. By sweeping over 200 threshold levels $\tau \in [0, 1]$, we obtained binary predictions at each threshold by classifying a window as “signal present” if the score exceeded τ . The true positive rate (TPR) and false positive rate (FPR) were defined as:

$$\text{TPR}(\tau) = \frac{\text{Correctly classified signal windows}}{\text{Total signal windows}}, \quad (2.6)$$

$$\text{FPR}(\tau) = \frac{\text{Incorrectly classified null windows}}{\text{Total null windows}}. \quad (2.7)$$

Varying the threshold τ traces out the ROC curve, revealing the sensitivity–specificity trade-off. To ensure robustness against stochastic variability in simulation and training, this procedure was repeated over 50 Monte Carlo runs, and the resulting TPR and FPR values were averaged. The **area under the ROC curve (AUC)** provides a scalar summary of performance, where values close to 1 indicate strong discriminative power, while values near 0.5 suggest random guessing.

Each mathematical model was simulated 50 times under different redness values (ranging from 0 to 0.5), and ROC–AUC values were calculated for all classifiers. The results,

summarized in Figure 2.5, compare the performance of the white-noise classifier (blue) and red-noise classifiers (orange) across six dynamical models. Each subplot corresponds to one mathematical model, with the y-axis showing the mean AUC score and the x-axis representing redness values. The figure demonstrates that, for each redness level, the classifier trained on white noise exhibits comparable performance to the classifier trained on red noise with the corresponding redness. While the overall performance remains high across models and noise conditions, a modest decline in AUC is observed for fold-type bifurcations, particularly in the Neural Activation and May’s Harvesting models, as the redness increases, indicating a slight sensitivity of these systems to stronger temporal correlations in the noise. This result confirms that the DL model trained on white noise generalizes effectively to time series containing coloured noise, maintaining high discriminative accuracy even under strong temporal correlations.

The results presented in this section demonstrate that our approach to developing a DL model for providing EWS of tipping points can be generalised to coloured noise and applied to highly coloured simulated or empirical time series. Moreover, these findings indicate that a classifier trained on white noise can still maintain a high level of performance on coloured time series.

2.5 Discussion

Conclusion

In this study, we developed a DL framework for detecting EWS of tipping points in complex dynamical systems. Our primary objective was to construct a training methodology grounded in theory by generating time series exclusively from the normal forms of canonical bifurcations. This approach eliminates the need for large, randomly sampled libraries of stochastic differential equations and ensures that the training data directly reflects the core dynamical patterns associated with critical transitions.

We first demonstrated that a CNN-LSTM model trained on these theoretically constructed time series is capable of accurately identifying EWS in out-of-sample data. This included both synthetic time series from unseen mathematical models and real-world empirical datasets, highlighting the model’s ability to generalise beyond the training distribution.

A second objective was to assess the robustness of this framework under realistic noise conditions. To this end, we replaced white noise in the training simulations with coloured (red) noise, introducing temporal autocorrelation using AR(1) processes. We found that the performance of the DL model remains stable across a wide range of redness values.

Finally, we performed a systematic evaluation using ROC curves and their corresponding

AUC scores. These experiments confirmed that models trained on coloured noise perform exceptionally well when tested on coloured stochastic time series with bifurcations. Importantly, we also showed that models trained on white noise generalise well to coloured noise settings, underscoring the flexibility of the training approach. Taken together, our results demonstrate that deep learning models trained on theoretically grounded simulations can serve as reliable tools for bifurcation detection, even in the presence of realistic stochasticity and noise correlations.

Limitation and Future Work

Nevertheless, the model and assumptions used in this study limited its scope. For instance, we focused exclusively on three specific types of local bifurcations. Global bifurcations could also be studied by expanding the training set to include these, as DL models can only recognise patterns they have been trained on. Furthermore, the study of coloured noise in our work was limited. Coloured noise refers to noise with a power spectral density (PSD) proportional to $1/f^\beta$, where f is frequency and β is the spectral exponent. Each value of β characterises a different type of noise. While our method creates a continuum between white and red noise ($\beta = 0$ and $\beta = 2$), it does not address other types of noise, such as blue noise ($\beta = -1$) and violet noise ($\beta = -2$). Moreover, the development of the training set relied entirely on normal forms, whereas incorporating higher-order terms could increase the accuracy of the DL model.

In future work, we aim to address these limitations. Specifically, we intend to expand the scope of this study to include other types of local and global bifurcations. Although climate empirical data predominantly rely on red noise, we plan to incorporate other types of coloured noise to create a more comprehensive classifier.

Chapter 3

Echoes Before Collapse: Deep Learning Detection of Flickering in Complex Systems¹

¹A preprint of this chapter is available on *arXiv* as: Babazadeh, Y.; Anand, M.; Bauch, C.T. [arXiv:2509.04683](https://arxiv.org/abs/2509.04683), posted 4 September 2025.

3.1 Abstract

Deep learning offers powerful tools for anticipating tipping points in complex systems, yet its potential for detecting flickering (noise-driven switching between coexisting stable states) remains unexplored. Flickering is a hallmark of reduced resilience in climate systems, ecosystems, financial markets, and other systems. It can precede critical regime shifts that are highly impactful but difficult to predict. Here we show that convolutional–long short-term memory (CNN–LSTM) models, trained on synthetic time series generated from simple polynomial functions with additive noise, can accurately identify flickering patterns. Despite being trained on simplified dynamics, our models generalize to diverse stochastic systems and reliably detect flickering in empirical datasets, including dormouse body temperature records and palaeoclimate proxies from the African Humid Period. These findings demonstrate that deep learning can extract early warning signals from noisy, nonlinear time series, providing a flexible framework for identifying instability across a wide range of dynamical systems.

3.2 Introduction

Sudden and irreversible shifts, tipping points [175, 65, 80, 136], can arise in systems as varied as climate [30, 122], ecosystems [66, 100, 155], financial markets [27, 108], and public health [209, 115, 12]. Anticipating such regime shifts is critical for mitigating their potentially severe consequences. Early warning signals (EWS) have emerged as an important tool for detecting the loss of resilience that precedes these critical transitions [200, 32, 42, 45, 57, 41, 44, 142, 25, 26, 116, 56, 111, 78].

One hallmark of reduced resilience is flickering: intermittent, noise-driven switching between coexisting stable states [45, 110, 9, 92]. Flickering has been documented in diverse systems, including palaeoclimate records [199, 195, 52, 53], neuroscience [101, 89], hibernation physiology [156], and lake ecosystem dynamics [207]. It often emerges before a bifurcation, when competing attractors become less stable and noise can drive rapid transitions between them [110, 53]. This makes flickering a valuable EWS that can complement indicators based on critical slowing down. However, existing approaches to detecting flickering rely mainly on statistical metrics such as variance, autocorrelation, and skewness [45]. These methods can be sensitive to noise characteristics, data resolution, and confounding processes, limiting their robustness in practice [45, 34].

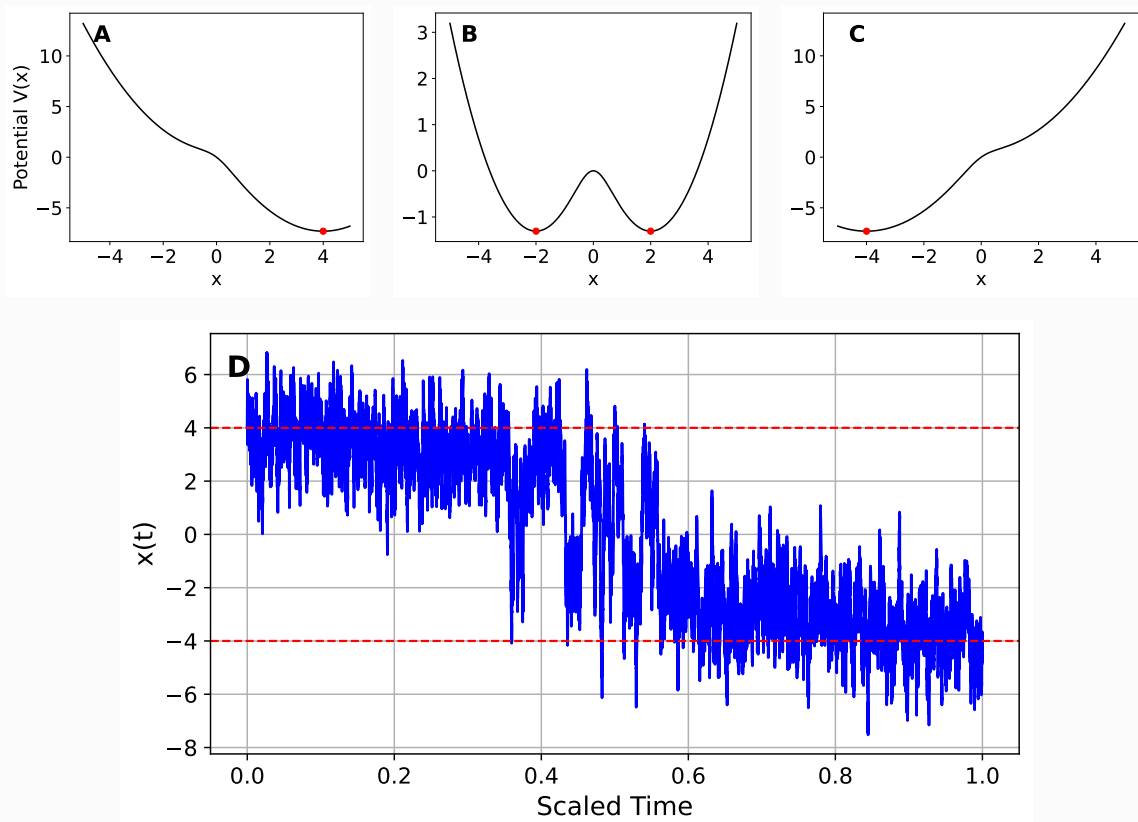
In parallel, advances in artificial intelligence (AI) and deep learning (DL) have demonstrated considerable potential for identifying EWS of tipping points, particularly for detecting bifurcations in noisy, nonlinear systems [34, 33, 159, 94, 46, 57, 59]. Yet, despite its

importance, the problem of detecting flickering using DL has not been addressed. This gap is significant: if DL models could reliably recognise flickering, they could expand our capacity to detect instability in a wider range of systems, especially when traditional statistical approaches fail.

Here, we develop a convolutional–long short-term memory (CNN–LSTM) framework to detect flickering patterns in noisy time series. We train the model on synthetic data generated from simple polynomial functions with additive noise, and test its generalization to a suite of more complex stochastic dynamical systems. We further show that the deep learning model consistently outperforms variance-based generic early warning indicators, achieving higher discrimination between true flickering and null cases. Finally, we apply the trained model to empirical datasets, including dormouse body temperature measurements [156] and palaeoclimate records from the African Humid Period [199]. Our results show that deep learning can robustly identify flickering across both simulated and real-world systems, offering a new and broadly applicable tool for anticipating critical transitions.

Box 1. What is flickering?

Flickering refers to noise-driven, intermittent transitions between coexisting stable states in a dynamical system. As a tipping point approaches, these switches become more frequent, indicating reduced resilience. (A–C) Potential landscapes $V(x)$ as the control parameter p changes, altering the number and stability of equilibria. In (B), two nearby equilibria coexist, and noise drives rapid switching between them. (D) Corresponding time series $x(t)$ with dashed lines marking the equilibrium positions from (A–C). The frequent switching in (B) is characteristic of flickering as an early warning signal.



3.3 Deep learning model and dataset construction

To detect flickering patterns in noisy time series (Box 1), we implemented a convolutional–long short-term memory (CNN–LSTM) network [177, 3, 88]. The CNN layers extract localized features across long temporal windows, while the LSTM layers capture temporal dependencies and preserve relevant sequence information through recurrent connections. The final dense layer outputs a binary classification indicating the presence or absence of flickering. This architecture was chosen for its proven ability to combine spatial feature extraction with long-range temporal modeling in time series analysis [58, 57, 33, 59, 34]. A schematic of the architecture is shown in Fig. 3.2, and all layer specifications are provided in the Appendix.

3.3.1 Details of Training Dataset

Synthetic time series were generated from a one-dimensional stochastic differential equation consisting of a linear term, a seventh-degree polynomial, and additive Gaussian noise:

$$\frac{dx(t)}{dt} = p - x(t) + [ax(t) + bx^2(t) + cx^3(t) + dx^4(t) + ex^5(t) + fx^6(t) + gx^7(t)] + \sigma\xi(t), \quad (3.1)$$

where $\xi(t)$ denotes Gaussian white noise and σ is the noise intensity.

The choice of a seventh-degree polynomial was made after systematic experimentation with polynomial orders ranging from three to seven. Lower-degree polynomials (e.g., cubic or quintic) were found to generate simpler potential landscapes which limited the diversity of dynamical behaviors available for training. In contrast, higher-order polynomials such as the seventh-degree form provide richer nonlinear structures capable of exhibiting multiple coexisting attractors and sharp transitions between them.

For the purpose of this research, we specifically required the polynomial term to exhibit a *saturational* behaviour, flattening for large $|x|$, to ensure the existence of at least two stable equilibria separated by one unstable equilibrium. This configuration naturally supports noise-induced switching, or flickering, between attractors (see Box 1), making it well-suited for building and testing the deep learning classifier with the desired accuracy and robustness across a variety of dynamic regimes.

Random Coefficient Generation.

The polynomial coefficients were randomly drawn from the following ranges to ensure diverse potential landscapes:

- $g \in (0, -2)$
- $f \in (-g, +g)$
- $e \in (0, -2)$
- $d \in (-e, +e)$
- $c \in (0, -2)$
- $b = 0$
- $a \in (1, 3)$

Initial Conditions and Noise.

We set $p = 5$, denoted as p_0 , to initialize the system well before the bifurcation point. This ensures that the dynamics begin in a single, stable attractor corresponding to non-flickering behaviour. Starting sufficiently far from the transition is essential for exposing the deep learning model not only to flickering dynamics but also to the gradual progression from non-flickering to flickering regimes as p approaches its critical value p^* (Fig 3.2).

The initial condition x_0 was chosen as the positive root of the right-hand side of Eq. (4), corresponding to the stable equilibrium of the potential well at p_0 . The noise intensity was set to $\sigma = 1.2x_0$. This choice balances the competing requirements of sufficient stochastic forcing and signal coherence. If the noise amplitude is too small, trajectories remain trapped within a single potential well and never reach the neighbouring basin of attraction, thereby preventing flickering. Conversely, if the noise amplitude is too large, the time series becomes dominated by stochastic fluctuations and loses the characteristic temporal structure required for classification. The optimal value of σ is therefore expected to scale with the width of the potential well, proportional to x_0 , and our experiments confirmed that a value of $1.2x_0$ yielded the most realistic flickering dynamics and best downstream model performance.

Determination of Transition Point.

The critical value p^* was determined by finding x^* where the slope of the polynomial term equals that of the linear term, that is, where the derivative of the nonlinear restoring force equals unity (slope = 1). At this point, the two curves, the linear term $x(t)$ and the nonlinear polynomial term, become tangent to each other, corresponding to the collision and annihilation of two equilibrium points. This geometric condition defines the

fold (saddle-node) bifurcation of the system, beyond which the upper stable equilibrium disappears and the system can only remain in the lower stable branch. In other words, the intersection of the linear and polynomial components represents the equilibria of the dynamical system, and the tangency condition marks the loss of bistability and the onset of a critical transition. The corresponding critical value p^* was then computed from:

$$y^* = x^* - p^*. \quad (3.2)$$

Simulation Procedure.

The stochastic differential equation was numerically integrated using the Euler–Maruyama method with a fixed time step of $\Delta t = 0.01$. Time series were generated with total lengths of 5000, 6000, 7000, 8000, 9000, and 10,000 steps. For flickering series, the control parameter p was linearly varied from its initial value p_0 to the critical value p^* , while for non-flickering series, p was kept constant at p_0 .

The rationale for generating time series of multiple lengths was to capture the inherent stochastic variability of flickering dynamics, which can manifest over different temporal scales. Flickering is not a single, sharply defined event but rather a stochastic process whose frequency and persistence evolve as the system approaches the bifurcation. Training the deep learning model on signals of varying durations therefore allows it to learn both short-term and long-term manifestations of this behaviour. The ensemble of models trained on different sequence lengths yields a cohesive representation capable of identifying the entire flickering window with high sensitivity.

We also found through numerical experimentation that, given our model configuration and parameter choices, the system must be simulated for at least 5000 integration steps to evolve smoothly from a distant non-flickering state ($p = p_0$) toward the flickering regime near p^* . Shorter runs did not provide sufficient temporal development for the transition dynamics to emerge clearly.

Dataset Composition.

A total of 2000 time series were generated for each category (flickering and non-flickering). For each series, we computed its rolling variance using a window size of 1000, and both the raw time series and the rolling variance (each followed by z-score normalisation) were used as the two input channels to the CNN–LSTM model. This dual-channel representation allows the network to jointly capture both the raw signal dynamics and the gradual variance escalation typical of flickering events. More details on the DL architecture are provided in Appendix B1.

Each CNN–LSTM model was trained on a specific sequence length from the set $\{5000, 6000, 7000, 8000, 9000, 10000\}$, allowing the ensemble to detect flickering patterns across multiple temporal scales. Shorter sequences capture rapid transitions, while longer ones resolve extended pre-transition dynamics. All models achieved final training accuracies above 95%, and ensemble averaging yielded robust detection across temporal scales.

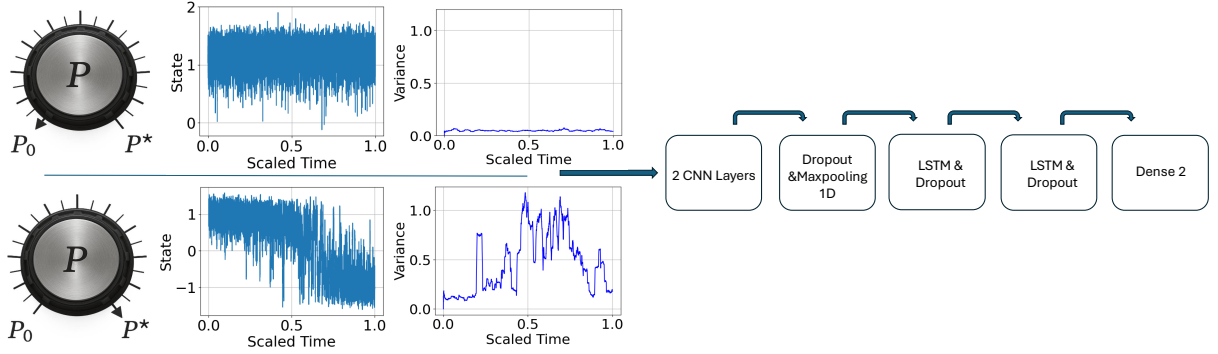
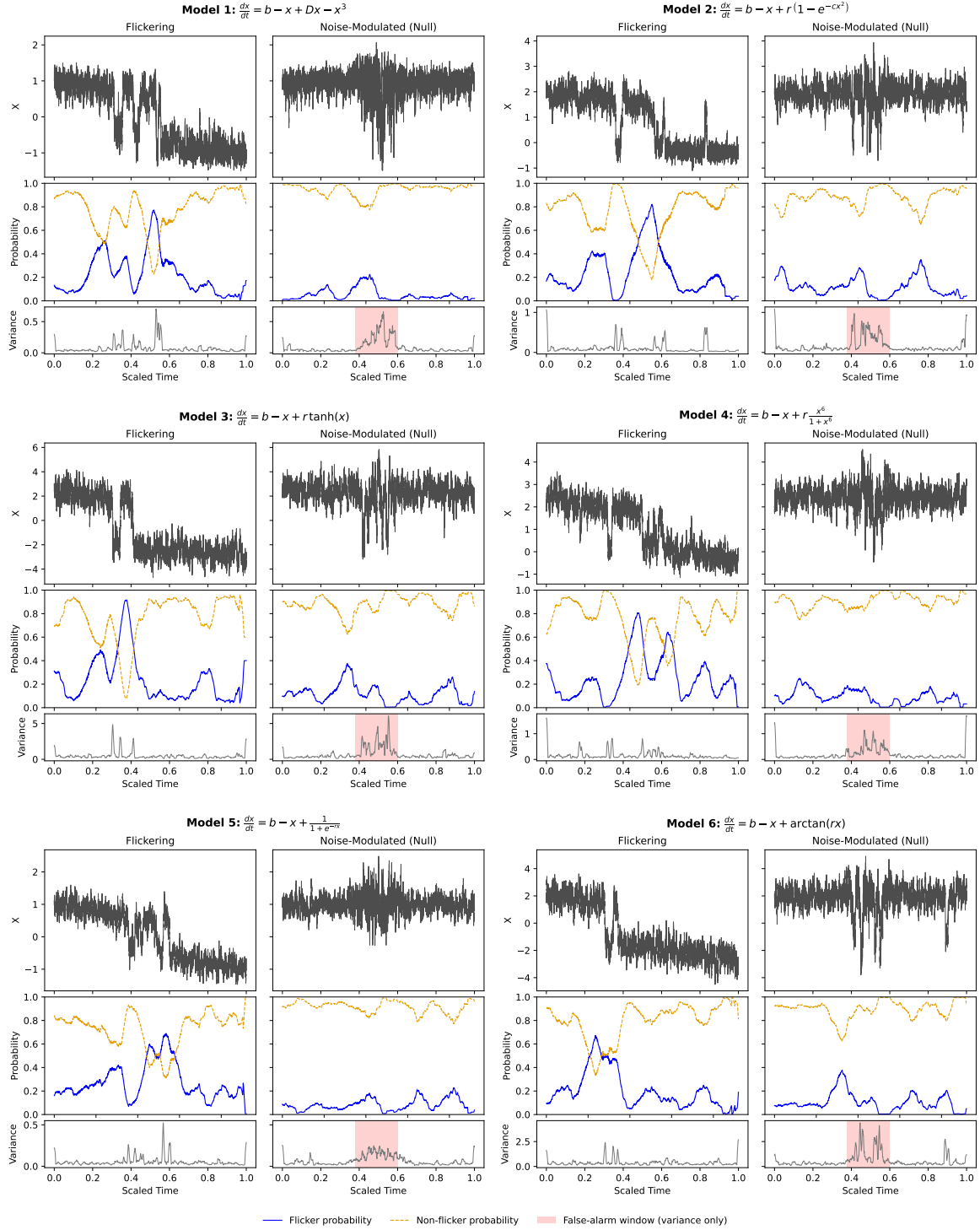


Figure 3.2: **Deep learning model architecture for flickering pattern detection.** The CNN–LSTM framework detects flickering patterns as early-warning indicators of tipping points. Simulated time series with and without flickering (parameterized by p_0 and p^*) are preprocessed (see Box 1) and used to train the model, which outputs a binary classification indicating the presence of flickering.

3.4 Results

Figure 3.3: **Deep-learning discrimination of true flickering versus noise-modulated null dynamics across six models.** For each model, the left column depicts genuine flickering between coexisting equilibria, while the right column shows a null trajectory where the noise amplitude is modulated to mimic flicker. Within each panel, the top trace is the state time series, the middle trace shows the classifier’s flicker (blue) and non-flicker (yellow) probabilities, and the bottom trace shows the moving variance. The red band marks a noise-driven interval used to probe false-alarm behavior in the null case; axes are aligned to enable direct visual comparison.

Deep-Learning Analysis of True Flickering vs Noise-Modulated (Null) Dynamics



3.4.1 Details of Simulated Models

To evaluate the generalization performance of the trained deep learning classifier, we tested it on six distinct nonlinear stochastic dynamical systems that were unseen during training. Each model represents a different class of one-dimensional stochastic differential equations (SDEs) exhibiting bistability and noise-driven switching under specific parameter regimes. For each system, the simulation was run for 62,500 time steps using the Euler–Maruyama method with a fixed time step of $\Delta t = 0.01$. All systems included an additive stochastic forcing term $\sigma\xi(t)$, where $\xi(t)$ denotes Gaussian white noise.

Two simulation regimes were implemented for each model: (1) **Flickering case:** a control parameter (b) was varied linearly over time from an initial to a final value, gradually reducing the stability of one attractor and inducing noise-driven switching between alternative states prior to the bifurcation point, while the noise amplitude σ was held constant. (2) **Non-flickering (null) case:** the control parameter b was held constant, but the noise amplitude σ was transiently increased in a triangular profile during the middle third of the simulation, producing variance inflation without actual state transitions.

This design ensured that both flickering and non-flickering time series exhibited comparable temporal variance dynamics, while differing fundamentally in their underlying attractor-switching behaviour. As a result, the classification task required the model to distinguish genuine approach-to-bifurcation dynamics from visually similar but mechanistically distinct variance changes which is a critical test of discriminatory power.

The six benchmark models and their corresponding parameter settings were as follows:

1. Cubic Polynomial Model:

$$\frac{dx}{dt} = b - x + Dx - x^3 + \sigma\xi(t), \quad (3.3)$$

with $D = 1.5$, $\sigma = 0.4$, and b linearly ramped from 0.5 to -1 in the flickering case.

2. Exponential Potential Model:

$$\frac{dx}{dt} = b - x + 2 \left(1 - e^{-2x^2}\right) + \sigma\xi(t), \quad (3.4)$$

with b decreasing from 0 to -1 in the flickering case and $\sigma = 0.45$.

3. Hyperbolic Tangent Model:

$$\frac{dx}{dt} = b - x + 2 \tanh(x) + \sigma\xi(t), \quad (3.5)$$

with b decreasing from 0.5 to -1 in the flickering case and $\sigma = 0.9$.

4. Sigmoidal Hill Function Model:

$$\frac{dx}{dt} = b - x + 1.5 \frac{x^6}{1 + x^6} + \sigma \xi(t), \quad (3.6)$$

with b decreasing from 1 to -0.5 in the flickering case and $\sigma = 0.5$.

5. Logistic Function Model:

$$\frac{dx}{dt} = b - x + \frac{1}{1 + e^{-10x}} + \sigma \xi(t), \quad (3.7)$$

with b decreasing from 0 to -1 in the flickering case and $\sigma = 0.3$.

6. Arctangent Potential Model:

$$\frac{dx}{dt} = b - x + \arctan(10x) + \sigma \xi(t), \quad (3.8)$$

with b decreasing from 0.5 to -1 in the flickering case and $\sigma = 0.8$.

3.4.2 Generalization Across Dynamical Systems

We next tested whether the classifier could reliably detect flickering and thereby provide an early-warning signal of an impending transition across this suite of six nonlinear one-dimensional systems that were structurally more complex than those used for training (Fig. 3.3; model equations shown above each panel). For each system, two classes of trajectories were generated: a *flickering condition* (left column), in which the control parameter was slowly advanced toward the tipping point to induce noise-driven switching between basins of attraction prior to the transition, and a *null condition* (right column), in which the control parameter was fixed while the noise amplitude was transiently increased and decreased to mimic variance inflation without genuine basin switching.

This null condition poses an intentionally difficult test: the rolling variance (bottom row of Fig. 3.3) increases in both scenarios and would ordinarily trigger a false positive for classical variance-based early-warning indicators. To overcome this, each time series was processed by an ensemble of CNN–LSTM detectors trained on different temporal contexts (window lengths equal to 8%, 9.6%, 11%, 13%, 14%, and 16% of the series length), and their posterior probabilities were averaged to obtain a single cohesive detection signal (middle row).

Across all six systems, the ensemble produced a sustained elevation of the “flicker” posterior that consistently anticipated the critical transition in the flickering condition,

while remaining near baseline in the null condition despite comparable variance inflation (Fig. 3.3, middle vs. bottom rows). In other words, when noise alone was modulated, the model correctly withheld an alarm even though conventional early-warning metrics would have falsely signalled during the shaded interval; by contrast, when genuine basin switching occurred, the model’s predicted probability rose sharply prior to the bifurcation. These results demonstrate that the classifier generalises effectively to novel dynamical systems not encountered during training, capturing higher-order dynamical signatures beyond moment-based variance growth, and thereby distinguishing true proximity to tipping points from spurious stochastic variance.

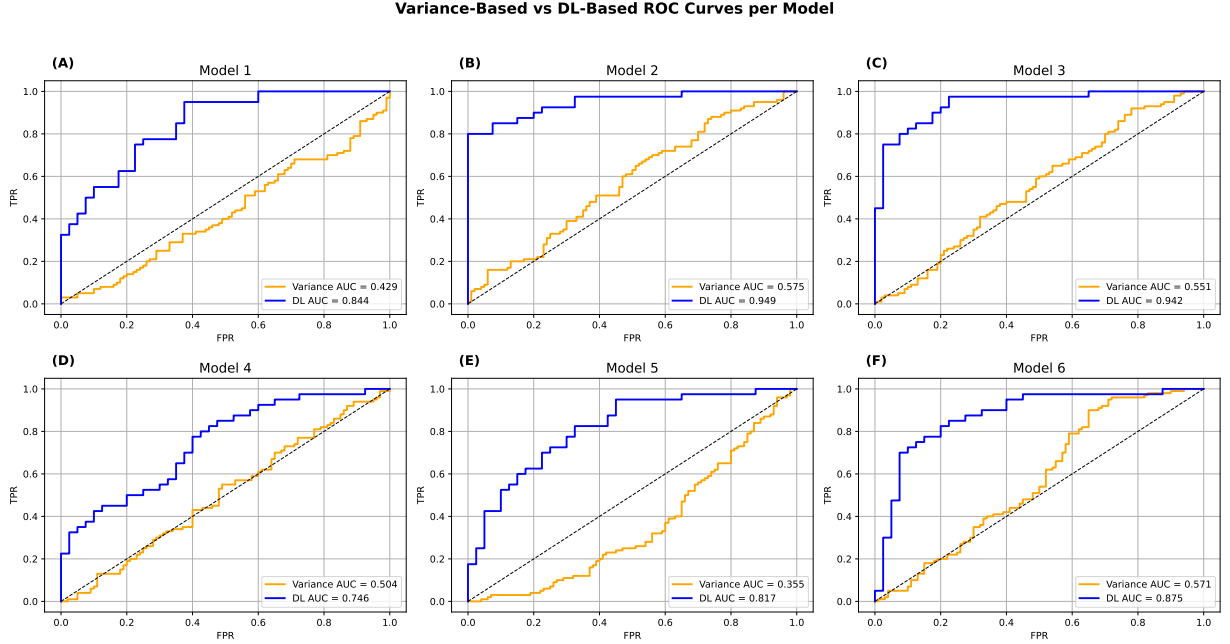


Figure 3.4: ROC performance on unseen dynamical systems. Receiver operating characteristic (ROC) curves comparing a variance-based indicator (orange) and the deep learning (DL) classifier (blue) for six nonlinear systems (A–F). Curves are based on 300 simulations per system under flickering and null conditions. Variance-based AUCs hover near 0.5, indicating poor discrimination when variance inflates in both cases, while DL-based curves are consistently bowed toward the upper-left, yielding markedly higher AUCs and capturing flickering-specific dynamics beyond variance growth.

3.4.3 Quantitative Evaluation via ROC Analysis

To quantitatively assess how well each method distinguishes genuine flickering from the challenging null condition described above, we computed receiver operating characteristic (ROC) curves for both the variance-based indicator and the deep learning (DL) classifier using the same ensemble of synthetic time series generated from the six unseen dynamical systems (Fig. 3.4A–F). In total, 300 independent stochastic simulations were performed under both flickering and null conditions, following the labelling protocol described in the preceding section. The ROC framework provides a threshold-independent measure of discriminative ability by comparing true- and false-positive rates across all possible classification thresholds. A classifier that performs no better than random yields a curve oscillating around the diagonal line ($\text{AUC} \approx 0.5$), whereas a perfect discriminator approaches the upper-left corner ($\text{AUC} = 1$).

ROC Curve Construction

For the DL detector, we generated 50 trajectories per dynamical model for each class. The DL detector consisted of an ensemble of CNN–LSTM networks trained on window lengths of 5,000, 6,000, 7,000, 8,000, 9,000, and 10,000 samples. For each trajectory, we computed the scalar score

$$s_{\text{DL}} = \max_w p_{\text{f}}(w) - \text{mean}_w[p_{\text{f}}(w)],$$

where $p_{\text{f}}(w)$ is the per-window flicker probability output by the DL model. To obtain a conservative estimate, the weakest score among the six DL models (i.e., the minimum s_{DL} across lengths) was retained for both flickering and null cases. The subtraction of the mean accounts for the normalization of p_{f} to $[0, 1]$.

For the variance-based baseline, we generated 100 trajectories per model for each class. We computed the rolling variance $V(t)$ with a window of 1,000 samples and defined the scalar score as

$$s_{\text{var}} = \frac{\max_t V(t)}{\text{mean}_t[V(t)] + \text{std}_t[V(t)]}.$$

The denominator normalizes the maximum variance surge by the baseline mean and variability, ensuring comparability despite the unscaled nature of $V(t)$.

In both cases, ROC curves were obtained by sweeping a classification threshold τ over the sorted set of scores $\{s_i\}$ and computing true- and false-positive rates at each τ . The area under the curve (AUC) was calculated using trapezoidal integration, and all analyses were performed separately for each dynamical model.

Results and Interpretation

Across all six systems, the variance-based indicator produced ROC curves that fluctuated near the diagonal ($AUC \approx 0.5$), confirming its inability to reliably distinguish flickering from null conditions when both exhibit similar variance inflation. In contrast, the DL classifier consistently yielded ROC curves that bowed sharply toward the upper-left corner, with substantially higher AUC values in every model tested (Fig. 3.4). This consistent improvement demonstrates that the deep learning detector identifies dynamical signatures characteristic of genuine flickering beyond simple variance growth and generalises effectively to structurally distinct systems unseen during training.

3.4.4 Application to Empirical Data

We further demonstrate that our deep learning approach can be extended to empirical records, enabling early detection of real-world tipping points through identification of flickering patterns.

Details of Empirical Data

The first dataset comprises high-resolution body temperature measurements from male and female edible dormice, recorded under Mediterranean outdoor conditions over the course of nearly one year. Temperature loggers were implanted to continuously measure hourly body temperature alongside ambient air temperature. The data exhibit two distinct regimes corresponding to active and hibernation phases, separated by abrupt transitions. During hibernation, both individuals displayed frequent and irregular alternations between torpid and euthermic states, indicative of flickering dynamics, that intensified prior to the transition back to activity. In contrast, body temperature remained relatively stable during active periods. Notably, the female entered hibernation approximately three weeks earlier than the male, although both resumed activity on the same day, suggesting both individual variation and synchronised physiological cues. These observations provide empirical support for interpreting flickering as an early warning of a critical physiological transition.

The second dataset is a high-resolution palaeoclimate record from the Chew Bahir basin in southern Ethiopia, spanning the termination of the African Humid Period (AHP) between 9 and 3 thousand years before present (kyr BP). This record is based on potassium concentrations in lake sediments, a well-established proxy for regional aridity [199]. The series reveals a sequence of approximately fourteen drought events, each lasting 20–80 years and recurring at intervals of roughly 160 years. These droughts precede and accompany a ~ 1000 -year transition from a stable wet to a stable dry climate state. The recurrence and

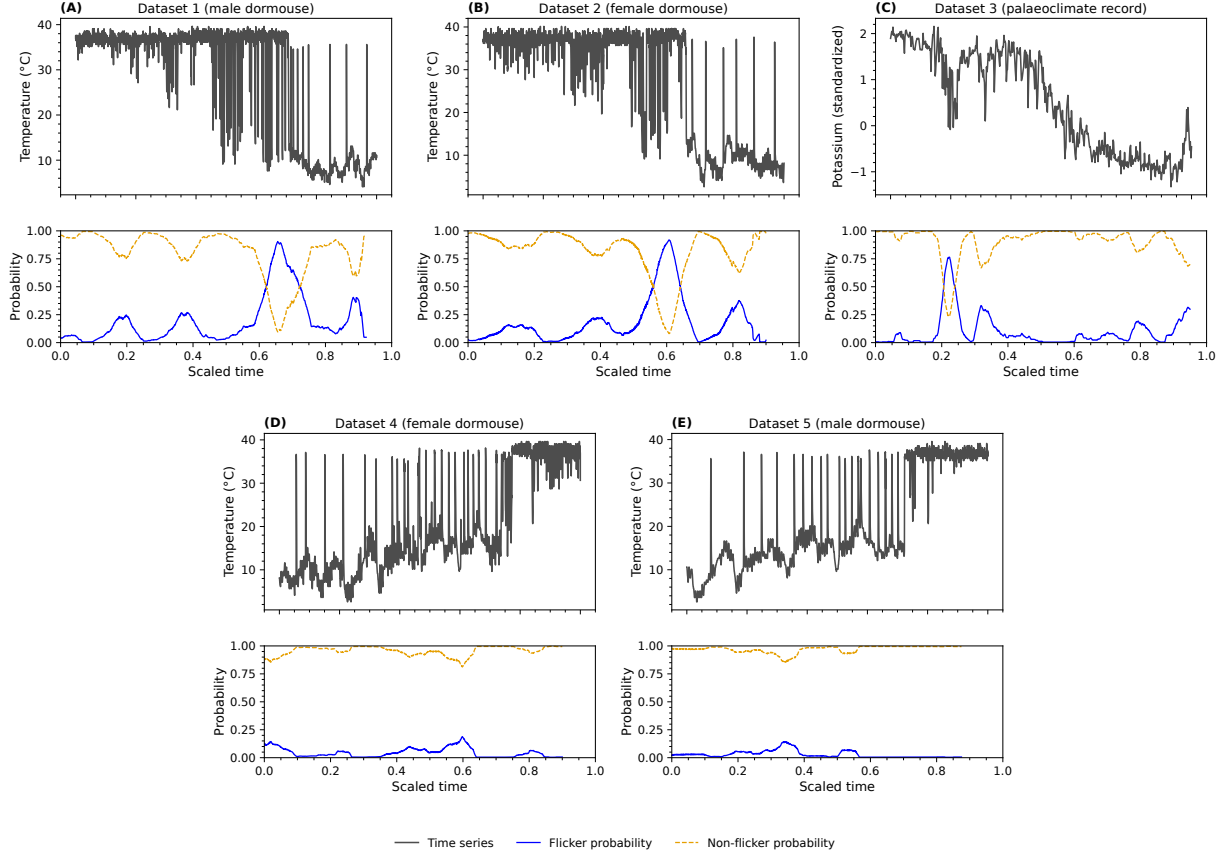


Figure 3.5: Application of the deep learning classifier to empirical datasets. (A,B) Body temperature records from male and female edible dormice during the transition from active to hibernation states. (C) Standardised potassium concentrations from the Chew Bahir palaeoclimate record in southern Ethiopia, a proxy for aridity, spanning 9–3 kyr BP and capturing the termination of the African Humid Period. In each panel, the upper subfigure shows the original time series (Grey); the lower subfigure shows the deep learning ensemble probabilities for flickering (blue) and non-flickering (orange). Clear pre-transition flickering signals are detected in (A–C). (D,E) Two additional dormouse temperature records containing flickering that the classifier failed to detect, highlighting cases of missed detection. Probabilities are obtained by averaging outputs from models trained with sliding windows spanning 8%, 9.6%, 11%, 13%, 14%, and 16% of the total time series length.

clustering of short drought and wet events generate a distinct pattern of climate flickering, rapid alternations between opposing climate states, that becomes increasingly pronounced as the system approaches the critical transition.

Model Application and Results

To ensure compatibility with the model input dimensions, all empirical time series were linearly interpolated to a uniform length of 100,000 time steps. Ensemble predictions were then obtained by averaging the outputs of CNN–LSTM models trained on sliding windows spanning 8%, 9.6%, 11%, 13%, 14%, and 16% of the total series length, consistent with the training configuration described earlier.

Across both datasets, the ensemble reliably identified flickering intervals preceding the observed transition points (Fig. 3.5A–C). In the dormouse data, clear flickering episodes were detected shortly before entry into hibernation, consistent with previous physiological analyses [156]. In the Chew Bahir palaeoclimate record, the classifier produced an elevated flicker probability during the centuries preceding the abrupt termination of the African Humid Period, capturing the dynamical signatures of pre-transition instability.

In contrast, two additional time series from the same dormouse dataset, where pre-transition fluctuations were less pronounced, did not elicit an elevated flicker probability (Fig. 3.5D–E). We speculate that the algorithm’s training set may need to include additional temporal features, such as the spacing and duration of intermediate temperature spikes, for it to detect subtler precursors in such cases.

Overall, these results illustrate both the applicability and the specificity of the approach: the model generalises successfully to biological and palaeoclimate data, generating early warning signals consistent with known transitions, while avoiding false positives when clear dynamical signatures of flickering are absent.

3.5 Discussion

We have demonstrated that deep learning models can provide reliable EWS of tipping points by detecting flickering patterns that often precede critical transitions. Our evaluation spanned three stages: (i) training on controlled synthetic datasets to capture essential flickering dynamics; (ii) testing on more complex stochastic models to assess generalisation; and (iii) applying the trained models to empirical datasets from biology and palaeoclimate. This pipeline demonstrates that deep learning can be trained on simplified representations yet retain predictive skill in diverse and realistic settings. Notably, the classifier maintained high performance even in challenging null scenarios where variance

inflation alone could mislead classical EWS, highlighting its ability to extract dynamical features beyond low-order statistical moments.

The framework has several limitations. First, not all critical transitions are preceded by flickering, and the method will not issue warnings in such cases. Second, the training set relied on relatively simple functional forms, polynomials up to the seventh degree, limiting exposure to more intricate or high-dimensional behaviours. Third, the analysis focused on univariate time series, whereas many real-world systems are multivariate and feature interacting subsystems. Addressing these limitations will broaden the scope and applicability of the method.

Future work will expand the diversity of the training set to include a broader range of functional forms, noise structures, and dynamical regimes. Integrating the deep learning framework with complementary EWS approaches, such as indicators of critical slowing down, may improve detection in cases where flickering is absent or weak. Extending the approach to multivariate signals will enable application to complex systems where multiple interacting variables influence stability, such as coupled climate–biosphere systems or physiological networks.

Overall, our results establish a generalisable framework for machine-learning-based detection of flickering that moves beyond traditional variance-based EWS. By bridging simple synthetic training data with complex, unseen, and empirical systems, this approach offers a scalable and adaptable pathway for anticipating critical transitions in domains ranging from climate risk assessment and ecosystem management to biomedical monitoring and financial stability analysis.

Chapter 4

Social dynamics can delay or prevent climate tipping points by speeding the adoption of climate change mitigation¹

¹This chapter has been peer reviewed and published as: Babazadeh, Y.; Anand, M.; Bauch, C.T. *Proceedings of the Royal Society A* **2025**, **481**(2322), 20250376. Publisher: The Royal Society. Publication date: 17 September 2025.

4.1 Abstract

Social dynamics are increasingly integrated into climate change models, yet tipping points in coupled social-climate systems remain understudied. This is concerning given that processes like permafrost thaw and forest dieback can trigger positive feedback loops that accelerate carbon release and global warming. We developed a coupled model combining an Earth system component with a social behavior model, each including tipping mechanisms. The climate model includes an additional carbon release term representing tipping reservoirs. The social model captures opinion dynamics shaped by social learning rates, mitigation costs, and the strength of social norms. Our results show that weak social norms have little effect on tipping dynamics. However, faster social learning can delay or even prevent climate tipping by accelerating mitigation responses. Simulations reveal nonlinear, regime-dependent dynamics, with bifurcation-like behavior separating tipping and no-tipping pathways. In some scenarios, a climate tipping point can trigger secondary tipping in social behavior. Crucially, the model enables estimation of the time to tipping, offering a forward-looking risk measure. These findings underscore the importance of incorporating social feedbacks in climate models to better anticipate, and potentially avoid, catastrophic tipping events.

4.2 Introduction

Using various measurement techniques, we know with confidence that the Earth’s average global temperature has increased by 1.1°C since 1800 due to anthropogenic activities [218]. For billions of years, Earth has developed intricate feedback mechanisms among various carbon reservoirs, balancing carbon emissions and absorption to achieve equilibrium. Examples of these feedback mechanisms include the ocean’s capacity to absorb excess atmospheric carbon and forests on land capturing carbon from the atmosphere [213] [82]. Since the beginning of industrialisation, excess human-caused carbon emissions have interfered with the function of these mechanisms [138]. In addition to direct carbon emissions, human activities can have indirect effects that can exacerbate the situation. Recent studies have revealed a reduced capacity of oceans and land to absorb atmospheric carbon due to ocean acidification and forest fires [190] [55]. This is among the causes of the rise in the global average temperature as more carbon is released into the atmosphere. Consequently, the ability of oceans and land to absorb carbon decreases, leaving more carbon in the atmosphere and forming a positive feedback loop. This positive feedback loop can increase the probability of irreversible events and a tipping point due to temperature rise [141], [61].

Multiple recent studies highlight the possible emergence of climate tipping points be-

yond a global temperature anomaly of 1.5°C [7, 126]. Several large carbon reservoirs may be triggered or accelerated by warming, including permafrost thaw, boreal forest dieback, Amazon rainforest dieback, and tropical peatland drying [145]. For example, rising temperatures cause permafrost in polar regions to thaw, releasing significant amounts of stored carbon as CO₂ and CH₄. Under the high-emissions RCP8.5 scenario, gradual thawing could release 23–174 GtC by 2100. This permafrost carbon feedback is projected to increase global temperatures by approximately 0.3–0.8°C by 2100 [206, 201, 149]. Additionally, rising global temperatures can irreversibly increase the frequency and severity of extreme events [121, 28]. For instance, between March 2023 and February 2024, forest fires released approximately 8.6 billion metric tons of CO₂ globally, a 16 percent increase compared to the two-decade average [103, 39]. Another example is Amazon deforestation, where crossing critical thresholds (e.g., 20–25 percent forest loss or 3–4°C warming) can trigger a shift to a degraded state, releasing large amounts of carbon and contributing 0.1°C to global warming over the next century [144]. These examples highlight additional positive feedback loops that can drive climate tipping points as temperatures rise, underscoring the need to incorporate them into Earth System Models (ESMs).

The primary way to combat global warming currently is mitigation through emissions reductions [160], [69], and based on the fraction of mitigators in a population, total carbon emissions can be controlled. In this view, Earth and people are seen as a coupled system, with people acting as elements of this generalised system [130], [140]. This underscores the importance of behaviour models in understanding the mitigation process. Many behaviour models have been developed to provide a better understanding of how a population forms and exchanges opinions [162] [1] [129] [220] [24].

In addition, in the past decades, there have been numerous efforts to develop a numerical structure for simulating climate models [170]. These models differ in the features and aspects that they include or exclude. Many of these models include behavioural models as well to offer a broad perspective of the future [216] [31] [147]. Additionally, studying the possibility and existence of tipping points in the climate system has received high attention in previous years and many papers have explored different facets [123] [124] [43] [34], including for climate tipping points [217].

Building on previous work [19, 30, 148, 187, 157, 16, 210, 211, 104], we aim to couple an Earth system model with a behavioural model that has room for feedback mechanisms that can lead to tipping points. We aim to develop a model to understand the effects of events able to create positive feedback loops and the impact of different mitigation strategies on controlling the damage of these events by simulating the coupled model. This model is not intended to serve as a highly realistic forecast of future climate conditions. Instead, it is designed as a simplified tool to explore hypotheses, test assumptions, and investigate potential dynamic patterns. While it offers valuable insights, it does not aim to replace

more comprehensive climate models.

We use the ESM and behavioural dynamics that were previously used to provide insight into the interaction between the climate system and social opinion formation [122] [31]. We introduce an additional carbon feedback term to represent the effect of runaway carbon emissions driven by positive feedback mechanisms. This term captures the potential for large-scale carbon release from vulnerable natural reservoirs, such as permafrost thaw, Boreal forest dieback, Amazon rainforest dieback, and tropical peatland drying [145]. As these processes are expected to intensify with rising global temperatures, we formulate this term to activate beyond a critical temperature threshold, simulating the onset of an irreversible (or difficult to reverse) climate tipping point. The formulation reflects both the non-linear escalation of carbon release and the heightened likelihood of tipping events under high-temperature scenarios. We also study the effect of social processes on the likelihood, strength, and timing of the climate tipping point. We show that the tipping point is preventable using specific mitigation strategies.

4.3 Model Description

4.3.1 Behaviour Dynamics

In this work, we use a behavioural dynamics model previously applied to social-climate systems [31][147]. Here, the population is assumed to be divided into two groups, mitigators and non-mitigators. Individuals in these two groups, decide whether to mitigate or not, based on the utility of each option. Although the utility of each option is affected by various factors such as political decisions, advertisements, social structures, economic indicators like inequality and GDP per capita, the severity of the climate state, corporate regulations, government policies, technological advancements, etc., recent studies have shown that public perceptions of climate change, the future cost and effectiveness of mitigation technologies, and the responsiveness of political institutions are important in explaining variation in emissions pathways and therefore the constraints on warming over the twenty-first century [150]. To simplify the complexity of social models without losing accuracy, many previous studies have used parameters that represent the same social factors [31] [147] [112]. Following the same approach in this work, we assume the utility of each option includes its associated cost, the costs associated with climate change, and the tendency to remain in the already chosen group.

The utility of being a mitigator is given by

$$e_M = -\alpha + c\hat{f}(T) + \delta x \quad (4.1)$$

Here, α is the cost of adopting mitigative strategies, The parameter c is a proportionality constant that determines the relative perceived cost of mitigation compared to non-mitigation. In the final form of the dynamics, c effectively scales how individuals interpret the cost of climate change impacts. Values $c > 1$ amplify the perceived cost, making mitigation appear more attractive relative to non-mitigation, whereas values $c < 1$ reduce the perceived cost, dampening the incentive to mitigate. Thus, c governs the strength of perceived climate-related costs within the social decision-making process and modulates how strongly individuals respond to environmental risk. $\hat{f}(T)$ is the cost of temperature anomaly of T degrees Celsius. δ is the strength of social norms and x is the fraction of mitigators of the population. The utility of a non-mitigator is given by:

$$e_N = -\gamma - \hat{f}(T) + \delta(1 - x) \quad (4.2)$$

where γ is the cost of non-mitigative behaviour, and its absolute value does not affect the dynamics since only the relative cost between mitigators and non-mitigators matters. For simplicity in our analysis, we assumed that both groups share the same value of δ . If we show the rate of change between these two groups (social learning rate) with κ then the total rate at which non-mitigators switch to being mitigators is:

$$r_{N \rightarrow M} = \kappa x(1 - x) \max(e_M - e_N, 0) \quad (4.3)$$

Also, the total rate at which the mitigators switch to being non-mitigators is:

$$r_{M \rightarrow N} = \kappa x(1 - x) \max(e_N - e_M, 0) \quad (4.4)$$

As a result, the net rate of change is:

$$\frac{dx}{dt} = r_{N \rightarrow M} - r_{M \rightarrow N} = \kappa x(1 - x)(e_M - e_N) \quad (4.5)$$

Using the terms we provided for e_M and e_N we will have:

$$\frac{dx}{dt} = \kappa x(1 - x)(\gamma - \alpha + (c + 1)\hat{f}(T) + \delta(2x - 1)) \quad (4.6)$$

which can be rewritten as:

$$\frac{dx}{dt} = \kappa x(1 - x)(-\beta + f(T) + \delta(2x - 1)) \quad (4.7)$$

Here $\beta = \alpha - \gamma$ which defines as net cost of mitigation and $f(T) = (c + 1)\hat{f}(T)$.

We assume a sigmoidal form for $f(T)$, since the perceived cost of global warming escalates rapidly beyond a critical temperature threshold [30].

$$\hat{f}(T) = \frac{\hat{f}_{max}}{1 + e^{-\omega(T-T_{crit})}} \quad (4.8)$$

Here f_{max} is the maximum cost, ω is the degree of nonlinearity of the sigmoid function with units of K^{-1} and T_{crit} is the critical temperature about which costs are most sensitive to change. we can combine c and $\hat{f}_{max}$:

$$f_{max} = (c + 1)\hat{f}_{max} \quad (4.9)$$

and rewrite the $f(T)$ as:

$$f(T) = \frac{f_{max}}{1 + e^{-\omega(T-T_{crit})}} \quad (4.10)$$

where the f_{max} is the maximum cost, ω is the degree of non-linearity of the sigmoid function, and T_{crit} is the critical temperature at which costs are most sensitive to change. The term $\delta(2x - 1)$ is a mechanistic representation of injunctive social norms [13] and has been derived from assumptions about majority-enforcing behaviour elsewhere [196].

4.4 Earth system model

To study the effect of social dynamics on climate change, we couple the social learning model with a reduced-form Earth system model (ESM) [122], combined with simplified ocean dynamics following [152]. Variants of this ESM have been widely used in climate-carbon studies or served as components of more complex models [30, 214]. It reliably reproduces historical temperature anomalies from pre-industrial levels [30].

4.4.1 Atmospheric carbon dynamics

The dynamics of atmospheric CO_2 are modified to include an anthropogenic emission term that depends on the proportion of non-mitigators:

$$\frac{dC_{at}}{dt} = \epsilon(t)(1 - x) - P + R_{veg} + R_{so} - F_{oc} + \frac{R_{max}}{1 + e^{R_0(T-T_c)}}, \quad (4.11)$$

where C_{at} is the deviation of atmospheric CO_2 from pre-industrial levels and $\epsilon(t)$ is the baseline emission rate in the absence of mitigation. P is photosynthetic carbon uptake, R_{veg} and R_{so} are plant and soil respiration fluxes, and F_{oc} is net ocean uptake.

The final term introduces a temperature-driven positive feedback representing the possibility of a carbon-cycle tipping point (see Section 4.2). We assume this term follows a sigmoidal form: $R_{\max}$ is the maximum additional carbon flux (GtCyr^{-1}), R_0 is the degree of nonlinearity of the response and has units of K^{-1} , and T_c is the critical temperature at which this term is triggered. Varying $R_{\max}$ allows us to represent a wide range of potential large-scale carbon releases, including permafrost thaw, boreal forest dieback, Amazon rainforest dieback, and tropical peatland drying. We refer to Eq. 4.11 as the *modified model* when this tipping term is included and as the *baseline model* when it is omitted.

4.4.2 Full Earth system model equations

The full coupled carbon–climate system consists of the following equations:

$$\frac{dC_{at}}{dt} = \epsilon(t)(1 - x) - P + R_{veg} + R_{so} - F_{oc} + \frac{R_{\max}}{1 + e^{R_0(T - T_c)}}, \quad (4.12)$$

$$\frac{dC_{oc}}{dt} = F_{oc}, \quad (4.13)$$

$$\frac{dC_{veg}}{dt} = P - R_{veg} - L, \quad (4.14)$$

$$\frac{dC_{so}}{dt} = L - R_{so}, \quad (4.15)$$

$$c \frac{dT}{dt} = (F_d - \sigma T^4) a_E. \quad (4.16)$$

Here C_{oc} , C_{veg} , and C_{so} are the ocean, vegetation, and soil carbon pools. Temperature T is the deviation in global surface temperature from pre-industrial values, c is the effective heat capacity of the climate system, σ is the Stefan–Boltzmann constant, and a_E is Earth’s emitting surface area.

4.4.3 Biogeochemical fluxes

Photosynthesis. Photosynthetic uptake follows:

$$P(C_{at}, T) = k_p C_{veg} k_{MM} \left(\frac{pCO_{2a} - k_c}{K_M + pCO_{2a} - k_c} \right) \left(\frac{(15 + T)^2 (25 - T)}{5625} \right), \quad (4.17)$$

valid for $pCO_{2a} \geq k_c$ and $-15 \leq T \leq 25$. Atmospheric CO_2 concentration (in moles) is

$$pCO_{2a} = \frac{f_{gtm}(C_{at} + C_{at0})}{k_a}, \quad (4.18)$$

where $f_{gtm} = 8.3259 \times 10^{13}$ is the conversion factor from GtC to moles.

Plant respiration.

$$R_{veg}(T, C_{veg}) = k_r C_{veg} k_A e^{-E_a/[R(T+T_0)]}. \quad (4.19)$$

Soil respiration.

$$R_{so}(T, C_{so}) = k_{sr} C_{so} k_B e^{-308.56/(T+T_0-227.13)}. \quad (4.20)$$

Turnover of vegetation.

$$L(C_{veg}) = k_t C_{veg}. \quad (4.21)$$

Ocean carbon flux.

$$F_{oc}(C_{at}, C_{oc}) = F_0 \chi \left(C_{at} - \zeta \frac{C_{at0}}{C_{oc0} C_{oc}} \right), \quad (4.22)$$

where χ is the characteristic solubility of CO_2 in water and ζ is the evasion factor.

4.4.4 Radiative forcing and temperature

Net absorbed surface radiation is given by

$$F_d = \frac{(1-A)S}{4} \left(1 + \frac{3\tau}{4} \right), \quad (4.23)$$

where A is surface albedo, S is solar flux, and τ is the greenhouse opacity, which is the term that introduces the carbon–climate coupling.

$$\tau(CO_2) = 1.73(pCO_2)^{0.263}, \quad (4.24)$$

$$\tau(H_2O) = 0.0126 (HP_0 e^{-L/(RT)})^{0.503}, \quad (4.25)$$

$$\tau(CH_4) = 0.0231. \quad (4.26)$$

Here H is relative humidity, P_0 the water vapour saturation constant, L is latent heat per mole of water, and R is the molar gas constant. All parameter values and functional forms are taken from [122].

4.4.5 Simulation

In this work, we simulate our model from year 1800 to 2200. We kept the behavioural dynamics inactive from 1800 to 2017, since any influence of social behaviour on carbon emissions during this historical period would already be reflected in the empirical data used to calibrate the model and for ϵ in this period, we used the values of the historical record [74]. For years after 2017, we projected the value of $\epsilon(t)$ using the following functional form:

$$\epsilon(t) = \epsilon(2017) + \frac{(t - 2017) \epsilon_{\max}}{(t - 2017 + s)}. \quad (4.27)$$

This formulation reflects that, due to limitations in resources and population growth, carbon emissions are expected to exhibit a saturational behaviour in the future rather than increase indefinitely [30].

Where $\epsilon_{\max}$ is the saturation value and s is the half-saturation constant. The initial values of all climate variables are zero since they represent the deviation from pre-industrial values. Social dynamics started in 2017 with a value of 5 percent for the initial fraction of mitigators (x_0). The values of climate parameters are drawn from the original Earth system model [122]. Social factors are more uncertain and have broader upper and lower limits. We applied the social parameter values reported in the original study to the behavioural model [31]. Based on the explanation in section 4.3.2, for the tipping term, we varied $R_{\max}$ across a broad range (0 to 5 GtC/year) to capture a wide spectrum of emission scenarios. For T_c we used two values of 2 and 3 degrees Celsius to define high- and low-risk cases. These thresholds were selected to represent realistic worst-case scenarios that may occur over the coming decades, corresponding to temperature levels at which several major carbon-cycle feedbacks (such as permafrost thaw or large-scale ecosystem dieback) are projected to become increasingly likely. For R_0 we used $5 K^{-1}$ to ensure a sharp transition to the tipping point. The values of all the used parameters are provided in the Appendix C3.

4.5 Results

Simulating different scenarios, we found that it is possible to diffuse or delay the climate tipping point through social actions. In this model, $R_{\max}$ (maximum value of the tipping term) represents the strength of the tipping term in the modified model, and κ (social learning rate), β (net cost of mitigation), and δ (strength of social norms) are parameters of the social behaviour model. Figure 4.1 illustrates simulations of modified and baseline models for two examples of κ and $R_{\max}$. Comparing the two plots, we observe that the

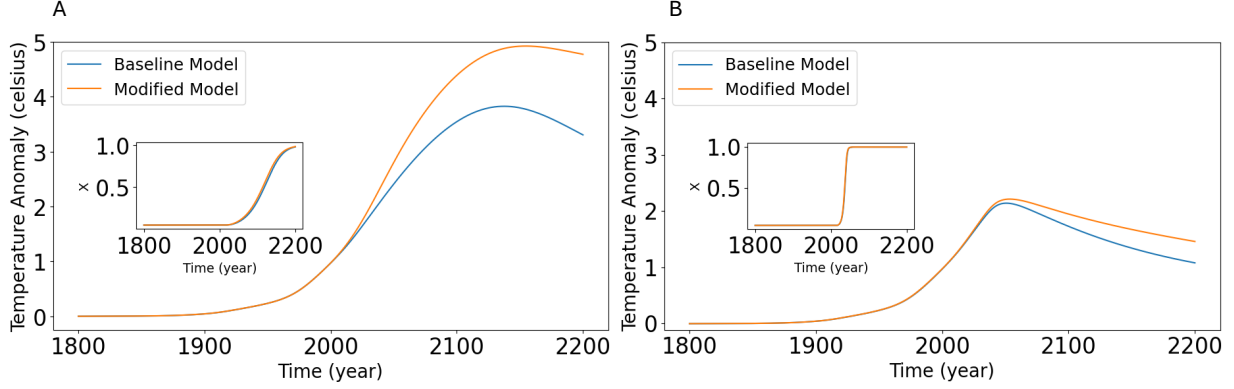


Figure 4.1: **Possibility of avoiding climate tipping point through social actions.** In this figure, The Baseline model is illustrated in blue and the modified model with an additional tipping term is illustrated in orange. The dynamic of x (fraction of mitigators) versus time is also depicted in both plots in a smaller box. Two plots differ in the values of κ and R_{max} (GtC/year) that were used: (0.01, 5) for plot A, where it shows tipping behaviour, and (0.1, 1) for plot B, where it does not.

tipping term can lead to a significant difference between the modified and baseline models for certain choices of κ and R_{max} , but not for others. This observation suggests that some social strategies can avoid the tipping point by deactivating the tipping term. To investigate and quantify the influence of the tipping term, we examine the area under the curve (AUC) of the global temperature anomaly time series, integrated from 1800 to 2200. Specifically, we compute the AUC for both the modified model (which includes the tipping term) and the baseline model (without the tipping term), and analyze the difference between them.

This AUC difference provides a compact and interpretable measure of the cumulative climatic impact introduced by the tipping mechanism. A larger difference suggests that the tipping feedback had a more pronounced effect, while a smaller difference may imply that social feedbacks effectively mitigated the influence of the tipping term. This metric serves as a diagnostic tool to systematically explore the sensitivity of climate outcomes to key social parameters, particularly the social learning rate, and helps reveal transitions between tipping and no-tipping regimes, consistent with bifurcation-like behavior in the coupled socio-climate system. Additionally, due to uncertainty about the value of critical temperature, we define two scenarios: 1) *high risk*, in which the critical temperature is set at 2°C, and 2) *low risk*, in which the critical temperature is set at 3°C.

Studying the difference in AUC versus social variables provides a more comprehensive insight into the potential to diffuse the tipping point. Figure 4.2 illustrates this analysis, revealing that the difference in AUC does not rely heavily on β (net cost of mitigation) and δ (strength of social norms). This outcome is expected, since both β (the net cost of mitigation) and δ (the strength of social norms) enter the baseline and modified dynamics in the same linear form. As a result, their contributions cancel out when computing the AUC difference between the two models, leading to minimal observable impact in our comparative sensitivity analysis.

However, our analysis depicts more complex behaviour when it comes to κ (social learning rate). As shown in panels A and D, isoclines fit into V-shaped regions, meaning that for each value of R_{max} , there exist two values for κ that result in the same difference in AUC, and there is one value of κ between them that maximises it. The tipping term is engaged inside these V-shaped regions, suggesting the possibility of a tipping point. A detailed sensitivity analysis of the impact of κ on climate outcomes is provided in the Appendix C2. We can also compare the results in two predefined scenarios of high ($T_c = 2$) and low risk ($T_c = 3$) in Figure 4.2. As shown, panels A, B, and C represent the high-risk scenario, depicting more concerning results and a higher possibility of tipping points due to more elevated differences in AUC in panels B and C, as well as larger V-shaped regions. On the other hand, panels D, E, and F show that in low-risk scenarios, the V-shaped regions are smaller, and the difference in AUC is less pronounced. Since κ has the greatest influence on the tipping term, we'll now focus primarily on results as a function of κ .

Social actions can also affect the time to the tipping point. In addition to analysing the evolution of temperature over time for the modified and baseline models, we are also interested in understanding the timing and how it is affected in the modified model under different mitigation strategies. First, we need to define the time to the tipping point. In this work, we measure the time when the temperature anomaly in the modified model becomes d times larger than the temperature in the baseline model at the same time, and refer to this specific instant as the *time to the tipping point*. We can now study the evolution of this time as a function of R_{max} and κ , which cover different scenarios of carbon emissions and mitigation strategies. Figure 4.3 illustrates this analysis. White regions represent cases where the tipping point did not occur. Coloured regions, representing cases with a tipping point, again fit into V-shaped areas similar to the results in 4.2. Results are provided for two predefined scenarios: high risk (plots A, B, and C) and low risk (plots D, E, and F). By comparing the worst cases of both scenarios, we can see that the tipping point occurs around the year 2120 in the low-risk scenario and around the year 2060 in the high-risk scenario. This suggests that the tipping point can happen 60 years sooner in the high-risk scenario. Additionally, in the low-risk scenario with a higher critical temperature, we observe that the V-shaped regions become smaller, meaning that fewer choices of R_{max}

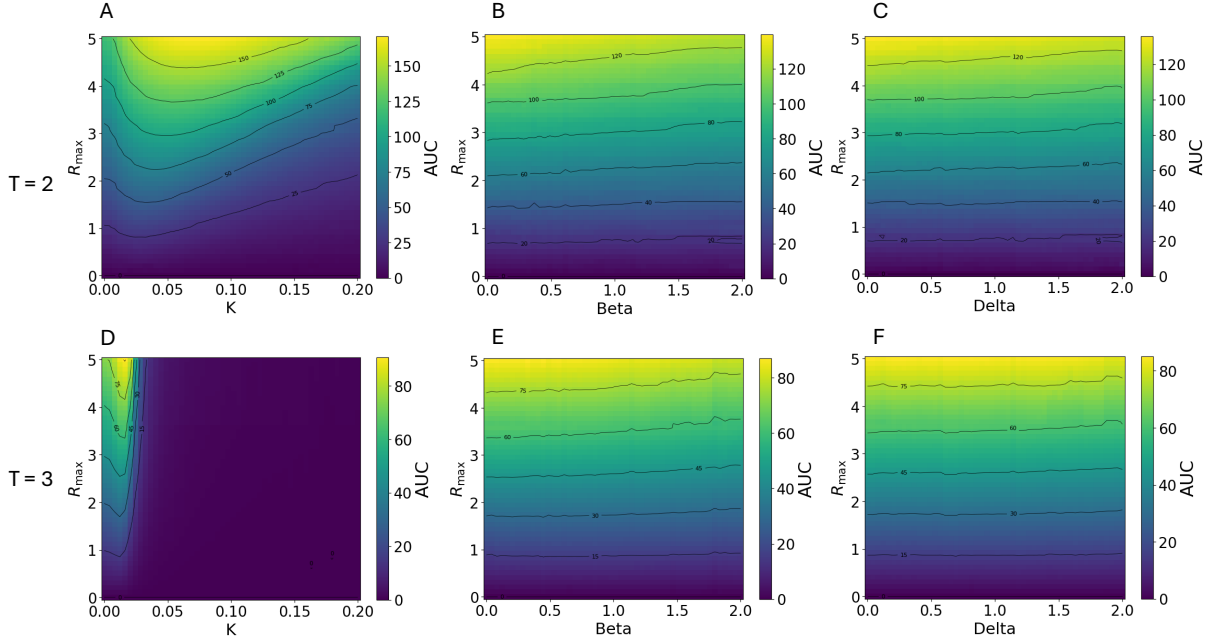


Figure 4.2: **The difference in AUC versus social behaviour variables.** In this figure, the difference in AUC is depicted against social behaviour parameters and R_{max} (GtC/year). The top plots represent the high-risk scenario ($T_c = 2$), and the bottom ones correspond to the low-risk scenario ($T_c = 3$). The plots illustrate the difference in AUC versus Social Learning Rate on the left, Net Cost of Mitigation in the middle, and Strength of Social Norm on the right.

and κ lead to a tipping point. We also observe that by selecting different sets of values for R_{max} and κ , the tipping point can be delayed by around 60 years in the low-risk scenario and 120 years in the high-risk scenario.

We have also shown that social interventions and the level of carbon emissions can influence the peak temperature deviation. Studying peak temperature deviation is essential as it provides valuable insights into the severity of each scenario and mitigation strategies. The analysis is depicted in Figure 4.4, where peak temperature anomaly is illustrated against R_{max} and κ for two predefined scenarios: high-risk (left) and low-risk (right). Both plots confirm that by increasing κ and decreasing R_{max} , peak temperature anomaly decreases. Additionally, comparing these two plots reveals that more severe cases tend to occur in high-risk scenarios due to their elevated values for peak temperature anomaly and bent isoclines, suggesting a slower decrease in peak temperature anomaly. This is an

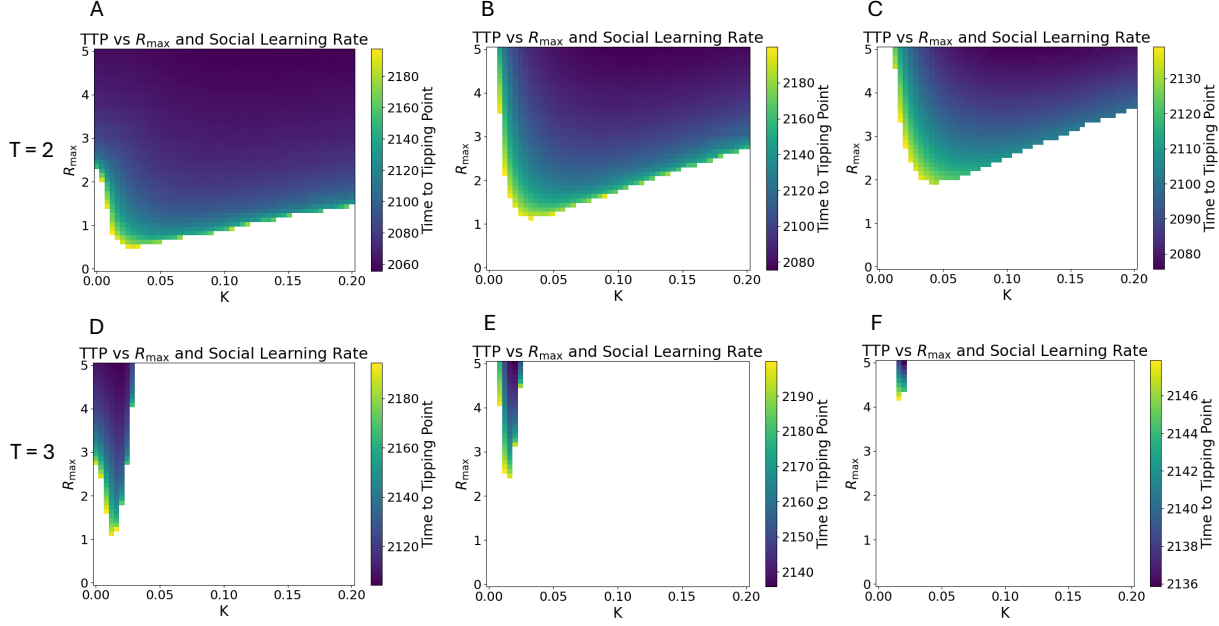


Figure 4.3: **Time to the tipping point for different emission and mitigation scenarios.** Time to the tipping point is plotted against $R_{\max}$ (GtC/year) and κ for six different cases. The top plots represent the results of the high-risk scenario ($T_c = 2$), and the bottom ones show the results of the low-risk scenario ($T_c = 3$). The plots on the left, middle, and right depict the results for different values of the threshold ($d = 1.1, 1.25$, and 1.5).

expected outcome based on the formulation of our model. In high-risk scenarios, where $R_{\max}$ is large, increases in κ can partially offset the resulting rise in peak temperature. In fact, due to the structure of the coupled socio-climate dynamics, the peak temperature anomaly can be approximately held constant by increasing κ and $R_{\max}$ simultaneously at an appropriate rate, which explains the pattern observed in the high-risk regime.

In addition, we have shown that critical temperature plays a pivotal role in determining the severity of our results and distinguishing between scenarios. So far, we have used two values of critical temperature to define high-risk and low-risk scenarios. Returning to the idea of the difference in AUC between the modified and baseline models, we examine the effect of critical temperature, varying it from 1.5 to 5 degrees while $R_{\max}$ is set to the constant value of 5 GtC/year. Figure 4.5 illustrates the outcome of this analysis. The difference in AUC against critical temperature and κ is depicted (left), and as shown, it

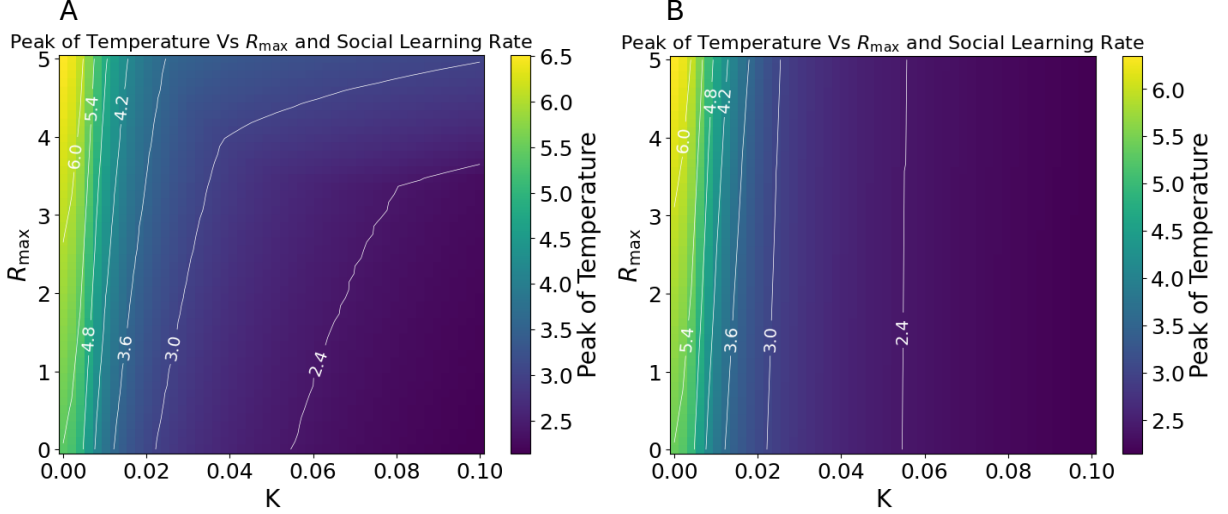


Figure 4.4: **Peak Temperature Anomaly for Different Emission and Mitigation Scenarios** Peak temperature anomaly (Celsius) is depicted against R_{max} (GtC/year) and κ for the high-risk scenario (left) and the low-risk scenario (right). Isoclines show that as κ increases, for the high-risk scenario, the decline in the peak of temperature is lower than low-risk scenario.

is split into two regions (yellow/green and blue). The yellow/green part represents the cases in which the difference in AUC is high and the tipping term is activated, causing the system to go through a tipping point, while the blue region represents cases where the system does not go through a tipping point. This clearly shows the bifurcated nature of our model. To provide deeper insights, two time series of temperature anomaly versus time are depicted for two examples. The orange circle represents the no-tipping case ($R_{max} = 3$ GtC/year, $\kappa = 0.05$), and the blue star represents the tipping case ($R_{max} = 1.6$ GtC/year, $\kappa = 0.09$). Another conclusion is that for each value of critical temperature, there is a threshold value for κ , beyond which the modified model does not go through a tipping point. For example, for $T_c = 3$, the tipping point only occurs if κ is greater than 0.03.

In addition to the analysis performed so far, the social model also exhibits a tipping point, and we have shown that the climate tipping point can trigger it. For high delta values, the social model can have up to three equilibrium points (at 0, 1, and one in (0,1)).

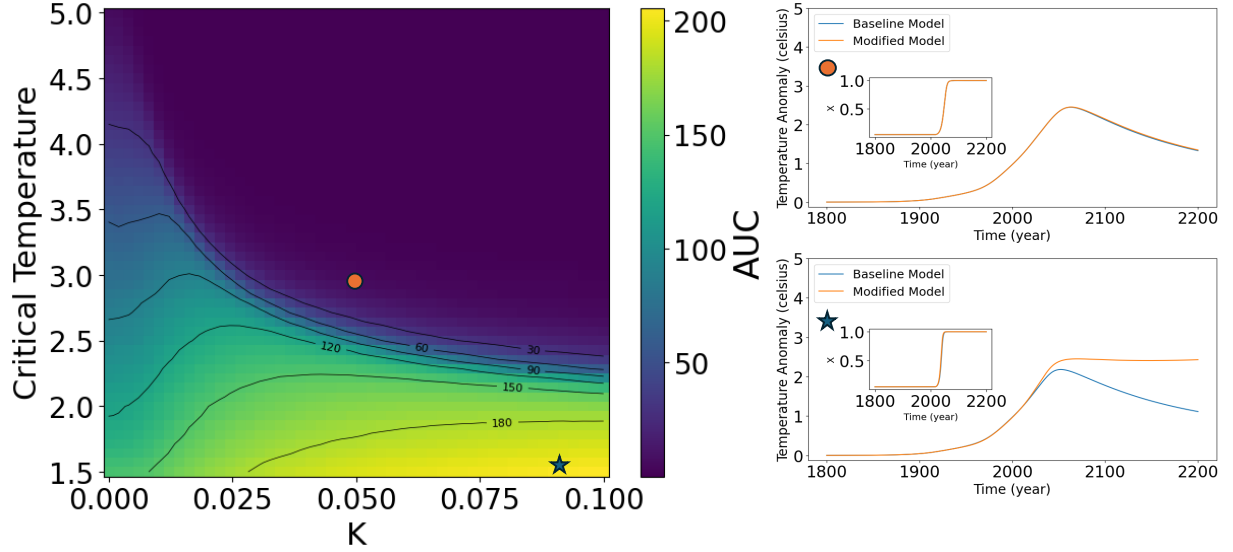


Figure 4.5: **The difference in AUC versus critical temperature and social learning rate** The Difference in Area under the curve (AUC) is depicted versus critical temperature and κ (left). Two examples representing two different choices of T_c and κ are chosen and illustrated on RHS. The plot at the top (orange circle) represents a case in which the modified model does not go through a tipping point, and the one at the bottom (blue star) represents a case of tipping point

To simplify the analysis, we define $-\beta + f(T)$ as ψ . The number and stability of these equilibrium points vary based on the value of ψ .

The social model has stable equilibrium points at 0 and 1, with an unstable point in between for high ψ , as shown in Figure 4.6. As ψ decreases, 0 becomes destabilised, leaving the system with a single stable equilibrium point at 1, as depicted in Figure 6(B). Since at the start of the simulations, x is very small and close to zero ($x_0 = 0.05$ in our study), in cases with high ψ , the social model converges to the stable equilibrium point at 0 (no mitigation). Conversely, when ψ is low, the social model converges to the only stable equilibrium point at 1 (full mitigation) (full stability analysis of the social model is provided in the Appendix C1).

As a result, for specific choices of climate parameters and for high delta ($\delta = 3$), when the climate model undergoes a tipping point and a sudden increase in temperature, $f(T)$

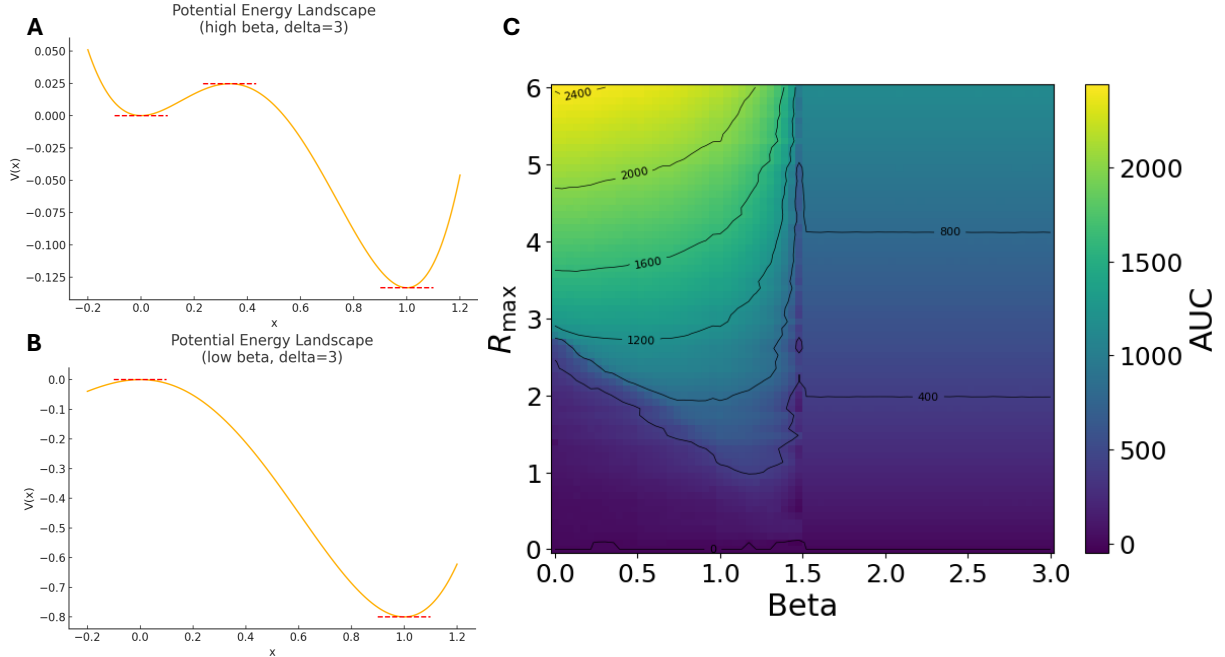


Figure 4.6: The Coupled Tipping Dynamics Between Climate and Social Models As shown in panels A and B, for high values of delta, decreasing beta destabilises 0 and triggers the social model to tip to $x=1$. A more detailed illustration of this effect in panel C shows that beyond a specific threshold value of beta, the tipping term does not lead to a significant difference between the modified and baseline model, whereas below that threshold, the tipping element activates.

increases, which raises ψ and destabilises 0, triggering the social model to tip to $x = 1$. The combination of the tipping points in the climate and social models results in an interesting dynamic, as shown in panel C. For high beta (net cost of mitigation), an increase in $f(T)$ during the climate tipping point does not cause ψ to increase sufficiently, and the social model converges to the stable equilibrium point at 0. Consequently, no dynamic interaction is observed. However, for smaller beta, ψ increases sufficiently due to the tipping point in the climate model, causing the social model to shift to the stable equilibrium point at 1, thereby generating diverse dynamics and interactions between the two models.

4.6 Discussion

If the Earth and its human populations qualify as a coupled social-climate system, then the concept of climate tipping points can be subsumed into the concept of social-climate tipping points. Our coupled social-climate model illustrates this change of framing and its impact on system predictions. For example, the model demonstrates how social dynamics can delay a climate tipping point indefinitely. Our analysis of social parameters revealed that the tipping point can be most effectively averted by accelerating the social learning rate. While other factors, such as the net cost of mitigation and the strength of social norms, influence the climate state, they were found to be largely ineffective in preventing tipping points, but we have also shown that for high values of social norms (δ), the tipping point in the climate model can trigger one in the social model. Additionally, the social learning rate was shown to significantly affect the time to tipping point. For each value of R_{max} , representing the strength of the tipping term, we identified a range of social learning rates that can avoid the tipping point and delay its onset. We also illustrated the effect of social learning on the peak of temperature anomalies.

To ensure a thorough analysis, we defined two scenarios based on the critical temperature that triggers the tipping term. It was shown that in high-risk scenarios, the climate state is more severely impacted, and a tipping point is more likely than in low-risk scenarios. Furthermore, we examined the effect of the critical temperature on the likelihood of a tipping point, finding that it plays a crucial role in distinguishing between tipping and no-tipping cases. Our results also reveal several non-obvious insights that go beyond the general intuition that mitigation helps avoid tipping points. We show that the coupled system exhibits bifurcation-like transitions between tipping and no-tipping regimes, highly sensitive to the social learning rate. Additionally, the inclusion of climate tipping can, in some cases, induce secondary tipping in social behavior. A sensitivity analysis further highlights the dominant role of social learning under tipping dynamics, reinforcing its importance as a key leverage point for policy interventions.

The model and assumptions we used limited the scope of our study. In the social behaviour model, we assumed a homogeneously mixed population, with decisions made individually by comparing the utility of mitigation versus non-mitigation. However, in reality, decisions occur at various societal levels—families, cities, and countries—and people’s choices are not binary but exist on a spectrum. Moreover, the logistic structure in our model assumes well-mixed interactions, which excludes important social features such as clustering, influence hierarchies, and local resistance. While this simplification supports analytical tractability, it may overlook dynamics that emerge in more realistic networked populations. Future extensions could address these limitations by incorporating spatial or network-based social structures. Additionally, the utility in our model is based on the

net cost of mitigation, the strength of social norms, and the cost of temperature anomalies. However, other factors, such as advertising and government policies, could also be relevant. Furthermore, human decision-making is not as straightforward as utility comparison; many psychological, social, and cultural factors play crucial roles, and these could be incorporated into future models to enhance realism. Moreover, we aimed to create a simplified model to generate hypotheses and explore new dynamic regimes. Consequently, the Earth system model and the additional tipping term we employed could be refined in future studies to produce more comprehensive insights and understanding.

In future work, we will focus on addressing the limitations of this study. Specifically, we are interested in exploring the similarities and differences in results by using modified and more complex social behaviour models. Additionally, the climate system can experience various types of bifurcations and tipping points such as: Fold bifurcation [127], Hopf bifurcation [165], Transcritical bifurcation [63], Fold-Hopf Codimension-2 Bifurcations [50]. Therefore, we aim to expand our study to include a broader range of bifurcations and examine their occurrence within an ESM model coupled with a social behaviour model. In addition, we plan to extend the current framework by incorporating stochastic tipping thresholds and multistable dynamics, to better capture the uncertainty and complexity of real-world tipping processes.

The findings of this study not only enhance our understanding of climate dynamics coupled with social behaviour but also highlight the need to accelerate mitigation efforts and the social learning rate to prevent tipping points. This research demonstrates the crucial role of mitigation in averting climate tipping points. Ultimately, we conclude it is vital to study social behaviour dynamics in the context of climate change and tipping points.

Chapter 5

When Simple is Enough: Binary Models Capture Social Complexity in Coupled Human–Environment Systems

5.1 Abstract

Models of coupled human–environment systems often face a tradeoff between realism and tractability. Spectrum opinion models, where social preferences vary continuously, offer descriptive richness but are computationally demanding and parameter-heavy. Binary formulations, in contrast, are analytically simpler but raise concerns about whether they can capture key socio–ecological feedbacks. Here we systematically compare binary and spectrum social models across four benchmark settings: (i) replicator dynamics coupled to a climate–carbon system, (ii) FJ opinion dynamics coupled to the climate–carbon system, (iii) replicator dynamics coupled to a forest–grassland ecological system, and (iv) FJ opinion dynamics coupled to a forest–grassland ecological system. We employ the relative integrated absolute error (RIAE) to quantify deviations between binary ($N = 2$) and spectrum ($N = 100$) formulations of social opinion dynamics in their feedback with ecological subsystems. Across systematic parameter sweeps of learning rates, reluctance, conformity, susceptibility, runaway amplitudes, and ecological turnover, the binary formulation typically tracks its spectrum counterpart to within $\leq 15\%$ for most parameter combinations. Deviations beyond this threshold arise primarily under very high social susceptibility or near-vanishing ecological turnover, where additional opinion modes and nonlinear feedbacks become consequential. We therefore present the binary formulation as a practical surrogate, not a universal replacement. As a rule of thumb, it is adequate when susceptibility is moderate, ecological turnover is appreciable, and runaway amplitudes are not extreme; in high-susceptibility or low-turnover regimes (especially near critical transitions), the full-spectrum model is preferable. This framing is intended to guide readers on when a binary reduction is sufficient versus when full-spectrum detail is warranted.

5.2 Introduction

Human societies and ecosystems are tightly intertwined, with feedbacks between social behavior and ecological processes shaping the trajectory of global environmental change [124, 128, 71, 154, 130, 22, 117]. Collective decisions about land use, resource exploitation, and climate mitigation can accelerate or prevent ecological tipping points, while ecological changes in turn reshape social preferences and behaviors [97, 83, 10, 182, 67]. In recent years, studies of coupled socio-ecological models have grown rapidly, reflecting a recognition that integrating human behavior into ecological modeling is essential for understanding resilience, forecasting critical transitions, and designing effective policy responses [66, 68, 188, 30, 19, 186, 18, 91, 184, 164].

A central challenge in socio-ecological modeling lies in how to represent social behavior.

Some models treat opinions or strategies as continuous variables, allowing individuals to occupy positions along a spectrum that captures nuanced attitudes and gradual shifts in preference [153, 49, 112, 172, 21, 4, 219, 23]. Others reduce social states to binary categories, such as cooperation versus defection, mitigation versus non-mitigation, or forest versus grassland preference [30, 10, 97, 189, 68, 15, 148, 194]. While spectrum models offer greater descriptive richness, binary models remain far more common because they are simpler to analyze, easier to parameterize, and computationally tractable [68]. Yet it remains unclear whether this simplification meaningfully alters system-level outcomes, or whether binary representations can faithfully approximate the dynamics of more complex spectrum models.

In this study, we address this gap by testing the robustness of binary social models against their spectrum counterparts in coupled socio-ecological systems. Specifically, we ask: Do binary models capture the essential dynamics of more complex spectrum social models? To answer this, we combine two ecological models with two distinct representations of social behavior, yielding four scenarios that allow systematic comparison across binary and spectrum formulations [10, 112, 97, 30]. This design enables us to assess whether the choice of social model meaningfully influences ecological outcomes, or whether binary models are sufficient approximations for capturing the critical feedbacks between human behavior and ecological dynamics.

Our results across four scenarios indicate that binary social models often approximate spectrum-model outcomes closely enough that the resulting system-level predictions remain effectively unchanged for modelling purposes, with relative errors typically below 15%. Our aim here is not to argue that binary models are universally best; rather, we provide guidance on when a binary reduction is sufficient. In many parameter regimes, such as moderate social susceptibility, appreciable ecological turnover, and away from runaway conditions, the binary formulation delivers substantial computational efficiency with little loss of explanatory power. Here, “runaway conditions” refer to parameter regimes in which strong positive environmental feedbacks drive rapid, self-amplifying ecological change. By contrast, under high susceptibility or very slow turnover, the full spectrum model is preferable.

These findings carry important implications for socio-ecological modeling. By validating the use of binary social models, our work supports the adoption of simpler formulations in large-scale or policy-oriented contexts where computational efficiency and interpretability are paramount. Binary models are easier to parameterize, more transparent in their assumptions, and more tractable for integration into complex ecological frameworks. Demonstrating that Binary Models approximate spectrum models well also helps unify ecological and social modeling approaches, offering a common foundation for exploring resilience, tipping points, and intervention strategies across a wide range of coupled human–environment

systems.

5.3 Methods

5.3.1 Forest–grassland mosaic model

As our first ecological model, we adopt the forest–grassland mosaic system developed by Innes et al [97]. In this framework, the ecosystem is composed of the proportion of land covered by forest $F(t)$ and grassland $G(t)$, where $F + G = 1$. The dynamics of forest cover are governed by recruitment of new trees into grassland and loss of forest through natural disturbances. This leads to the differential equation

$$\frac{dF}{dt} = w(F)F(1 - F) - vF, \quad (5.1)$$

where v is the rate at which forest reverts to grassland due to natural disturbances, and $w(F)$ is a recruitment function that depends on forest cover. The recruitment term $w(F)F(1 - F)$ captures the density-dependent establishment of new forest: recruitment requires both existing forest (seed sources) and available grassland (space for colonization).

The function $w(F)$ accounts for the mediating role of fire, which strongly suppresses recruitment when forest cover is sparse, but has weaker effects once dense stands of trees are established. This is represented by a sigmoidal function,

$$w(F) = \frac{c}{1 + e^{-k\frac{F}{1-F} + b}}, \quad (5.2)$$

where c controls the maximum recruitment rate, b sets the baseline recruitment at low forest cover, and k controls the abruptness of the transition between low and high recruitment regimes. This model yields multiple stable states: a grassland-dominated equilibrium ($F^* = 0$) and an interior equilibrium ($F^* > 0$) where both forest and grassland coexist, depending on parameter values and initial conditions.

The forest–grassland mosaic dynamics can be extended to include human influence through a harvesting function $J(x)$, where x represents the indicator of social opinion, capturing the fraction of individuals supporting land-use conversion. This function reflects the net effect of collective social preferences on the rate of forest harvesting or grassland expansion. . The coupled ecological equation becomes

$$\frac{dF}{dt} = w(F)F(1 - F) - vF + J(x), \quad (5.3)$$

where $J(x)$ captures the contribution of social opinion to deforestation or reforestation.

5.3.2 Climate–carbon model

For our second ecological model, we consider a reduced Earth system model with four carbon pools and an energy balance for global-mean temperature anomaly [122, 152, 10, 30]. Let $C_a(t)$, $C_{oc}(t)$, $C_v(t)$, and $C_{so}(t)$ denote, respectively, atmospheric, oceanic, vegetation, and soil carbon anomalies relative to baseline stocks $(C_{a0}, C_{oc0}, C_{v0}, C_{so0})$. Let $T(t)$ be the temperature anomaly (in °C) relative to a baseline T_0 , the dynamics are

$$\frac{dC_{at}}{dt} = \epsilon(t) - P + R_{veg} + R_{so} - F_{oc} \quad (5.4)$$

$$\frac{dC_{oc}}{dt} = F_{oc} \quad (5.5)$$

$$\frac{dC_{veg}}{dt} = P - R_{veg} - L \quad (5.6)$$

$$\frac{dC_{so}}{dt} = L - R_{so} \quad (5.7)$$

$$c \frac{dT}{dt} = (F_d - \sigma T^4) a_E \quad (5.8)$$

where $\epsilon(t)$ is anthropogenic emissions, a_E is Earth’s emitting area, c is the effective heat capacity of the climate system, and σ is the Stefan–Boltzmann constant. This is the same ESM model used in chapter 4. Full details of this model, including the definitions of all functions and parameters, are provided in the Appendix D1. Exogenous gross emissions $E(t)$ are reduced by mitigation behavior according to a time-varying mean effort $\bar{m}(t)$ defined in the social layer. In general, $\bar{m}(t)$ can be defined over different ranges depending on the formulation of the social model. For example, if $\bar{m}(t) \in [-1, 1]$, then effective emissions are given by

$$\epsilon(t) = E(t) \frac{1}{2} (1 - \bar{m}(t)), \quad (5.9)$$

so that $\bar{m}(t) = -1$ corresponds to no mitigation ($\epsilon = E$) and $\bar{m}(t) = 1$ corresponds to full mitigation ($\epsilon = 0$).

5.3.3 Replicator social model

As our first social model, we employ a replicator dynamics formulation in which the population is partitioned into N belief categories [10, 30]. Let $x_i(t)$ denote the fraction of the population in category $i = 0, \dots, N-1$, with $\sum_{i=0}^{N-1} x_i = 1$. Each category is associated with a belief value

$$m_i = -1 + \frac{2i}{N-1}, \quad i = 0, \dots, N-1,$$

so that opinions range continuously from -1 , indicating a strong grassland preference or low climate mitigation, to $+1$, indicating a strong forest preference or high climate mitigation.

Each category has an associated utility U_i and the mean utility in the population is $\bar{U} = \sum_i x_i U_i(F)$. Social imitation dynamics are then given by the replicator equation,

$$\frac{dx_i}{dt} = \kappa x_i (U_i(F) - \bar{U}),$$

where κ is the social learning rate, categories with above-average utility grow in prevalence, while those with below-average utility decline. The replicator social model is coupled to the ecological dynamics by including the mean belief

$$\langle m \rangle = \sum_{i=0}^{N-1} m_i x_i,$$

This framework naturally interpolates between binary and continuous opinion models by varying N .

- For $N = 2$, the opinion space collapses to two categories, $m_0 = -1$ and $m_1 = +1$. This corresponds to a binary social model, in which individuals can only prefer either grassland (equivalently, climate mitigation) or forest (equivalently, climate non-mitigation). The replicator equation then reduces to a two-state competition between these discrete preferences.
- For large N (In this work $N = 100$), the categories m_i densely cover the interval $[-1, 1]$, yielding a continuous spectrum of opinions. In this case the replicator dynamics approximate an opinion distribution evolving smoothly in response to ecological feedback.

In this way, the replicator formulation provides a unifying framework that can generate either binary or continuous social dynamics depending on the choice of N .

5.3.4 FJ social model

We model opinion dynamics with a Friedkin–Johnsen–type process on a bounded, scalar opinion $X_i(t) \in [-1, 1]$ for each agent $i = 1, \dots, n$. Let x_i^0 denote the agent’s private anchor, and let $\lambda_i \in [0, 1]$ be the agent’s susceptibility to social influence (heterogeneous, with values sampled from a distribution centered at λ_0 and bounded within $[0, 1]$) [112, 153].

$$\frac{dX_i}{dt} = \Psi \left[\underbrace{\lambda_i \left(\sum_{j \neq i} w_{ij} (X_j - X_i) + R \right)}_{\text{social influence + climate forcing}} + \underbrace{(1 - \lambda_i)(x_i^0 - X_i)}_{\text{anchoring to private belief}} \right], \quad X_i \in [-1, 1], \quad (5.10)$$

The pairwise weights are distance-decaying in opinion space,

$$w_{ij} = \exp\left(-\frac{|X_i - X_j|}{A}\right), \quad w_{ii} = 0, \quad (5.11)$$

where A is the sensitivity of interaction strength to the opinion distance between individuals, and R represents an exogenous runaway driver that shifts opinions in response to changes in the system and Ψ is an overall social adjustment rate.

This dynamical system (5.10) produces a spectrum model when we use the raw opinions $X_i \in [-1, 1]$ and average them,

$$\bar{X}_{\text{spec}}(t) = \frac{1}{n} \sum_{i=1}^n X_i(t),$$

To obtain a binary version that retains only group polarity, we map each opinion to its sign and then average. Values equal to zero are assigned to $+1$ or -1 with equal probability:

$$\tilde{X}_i(t) = \begin{cases} +1, & X_i(t) > 0, \\ -1, & X_i(t) < 0, \\ \text{randomly } \pm 1, & X_i(t) = 0, \end{cases} \quad \bar{X}_{\text{bin}}(t) = \frac{1}{n} \sum_{i=1}^n \tilde{X}_i(t).$$

5.3.5 Comparison Metric

To evaluate whether binary and continuous social models produce comparable outcomes when coupled to the same ecological dynamics, we require a metric that captures the difference between two time series in a way that is both interpretable and scale-free. We employ the relative integrated absolute error (RIAE), which is an L^1 absolute relative gap measure [171].

Formally, given two trajectories $f_1(t)$ and $f_2(t)$ defined on a time interval $[t_0, T]$, the RIAE is

$$\text{RIAE}(f_1, f_2) = \frac{\int_{t_0}^T |f_1(t) - f_2(t)| dt}{\int_{t_0}^T |f_{\text{ref}}(t)| dt} \quad (5.12)$$

where f_{ref} is a chosen reference trajectory (typically one of the two models). The numerator measures the integrated L^1 distance between the two trajectories, while the denominator normalizes by the total magnitude of the reference series. This normalization makes the RIAE unitless and directly interpretable as a percentage difference.

In all simulations, the social model is initially inactive to allow the ecological variables to evolve without social feedback. We then activate the social dynamics at time t_0 , and only from this point do we compute the RIAE. The integration therefore spans the interval from social onset until either (i) the system approaches its new equilibrium state or (ii) a specified maximum evaluation time is reached, depending on the scenario. This ensures that the comparison focuses exclusively on the time window when binary versus continuous opinion dynamics can exert influence on the ecological model 5.1.

This metric has several desirable properties for our setting. First, it is sensitive to cumulative differences across time, rather than to isolated pointwise deviations, thereby accounting for the entire ecological trajectory over the chosen evaluation window. Second, because it is normalized, it allows comparison across different ecological variables (e.g. temperature, forest cover, or mean opinion) that may have different scales and units. Third, expressing the RIAE as a percentage facilitates interpretation: for example, $\text{RIAE} = 0.10$ indicates that the binary model deviates from the spectrum model by an average of 10% of the reference magnitude over the evaluation period.

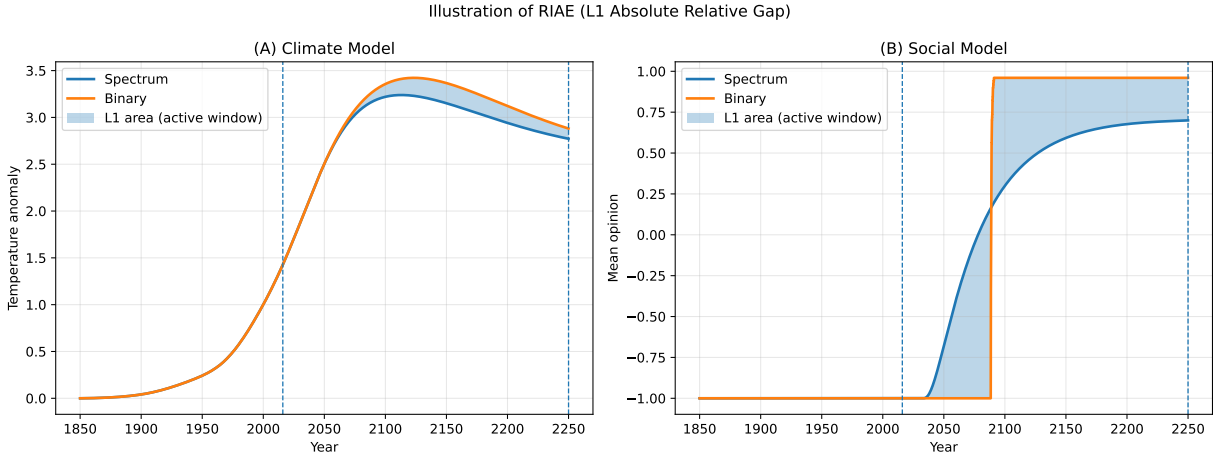


Figure 5.1: Illustration of the relative integrated absolute error (RIAE) metric. (A) Climate dynamics: comparison between spectrum and binary social models, with the shaded region indicating the integrated absolute difference after the onset of social dynamics (t_0) until the end of the evaluation window. (B) Social dynamics: comparison of mean opinion trajectories under spectrum and binary formulations, with shaded areas likewise representing the integrated L^1 gap.

5.4 Results

5.4.1 Coupling Between the Replicator Social Model and the Climate–Carbon Model

The coupling between the replicator social dynamics and the climate–carbon model operates through two feedback channels, linking temperature anomaly to social utilities and mean social opinion to emissions.

The temperature anomaly $T(t)$ enters the replicator dynamics by shaping the utility of each opinion category. Specifically, the utility assigned to category i is

$$U_i(T) = -\alpha_i + \beta_i f(T) + \delta x_i, \quad (5.13)$$

where α_i is a baseline reluctance parameter. The values of α_i are assigned along a linear grid from 0 up to $\alpha_{\max}$, with the number of steps equal to the number of opinion categories N . β_i is the mitigation effort associated with category i that goes from 0 to $\beta_{\max}$ similar to α_i , and δ is a conformity term. For simplicity of analysis, we assume that δ is the same for all groups, consistent with the modelling assumptions used in chapter 4 and [30]. The parameter $\beta_{\max}$ controls the amplitude or strength of the perceived climate cost, scaling how strongly $f(T)$ influences the utility of each category.

$$f(T) = \frac{f_{\max}}{1 + \exp[-\omega(T - T_c)]} \quad (5.14)$$

increases with temperature anomaly and represents the perceived benefit of mitigation under ecological stress. In this way, rising T increases the utility of higher-mitigation categories, biasing imitation dynamics toward stronger mitigation.

Conversely, the social state influences the climate system by reducing anthropogenic emissions. The population distribution $x(t)$ determines the mean mitigation effort

$$\bar{m}(t) = \sum_i m_i x_i(t), \quad (5.15)$$

which rescales gross exogenous emissions $E(t)$ to yield effective emissions,

$$\varepsilon(t) = E(t) \frac{(1 - \bar{m}(t))}{2}. \quad (5.16)$$

These effective emissions $\varepsilon(t)$ drive the atmospheric carbon equation in the climate–carbon model.

For $E(t)$ we used empirical emissions data up to the year 2017 [74]. Beyond this point, future emissions were modeled with a saturating function of time [30]:

$$E(t) = E(2017) + \frac{(t - 2017) \varepsilon_{\max}}{t - 2017 + s}, \quad t > 2017, \quad (5.17)$$

where $\varepsilon_{\max}$ is the maximum additional emissions rate and s is a saturation constant. This functional form produces a gradual increase that asymptotically approaches a finite level, preventing unrealistic divergence of future emissions.

To systematically compare the binary and spectrum replicator social models in the coupled climate-carbon system, we performed parameter sweeps across four key social parameters ($\kappa, \alpha, \beta, \delta$). In each case, we varied one parameter across a prescribed range in 10 steps, while keeping the others fixed at baseline values that corresponds a moderate social learning rate ($\kappa = 0.05$), a moderate maximum reluctance parameter ($\alpha_{\max} = 1$), unit social conformity strength ($\delta = 1$), and unit maximum mitigation effort ($m_{\max} = 1$). For each sweep, we considered two values of the critical temperature threshold, $T_c \in \{2.0, 3.0\}$, representing low and high climate sensitivity scenarios.

The parameter ranges explored were:

- κ : 0.01 to 0.10,
- $\alpha_{\max}$: 0.0 to 2.0
- $\beta_{\max}$: 0.0 to 2.0
- δ : 0.0 to 2.0

For each parameter setting, we simulated both the binary ($N = 2$) and spectrum ($N = 100$) replicator models under otherwise identical ecological conditions. The coupled model was run from the year 1800 to 2200, with the social model kept inactive until 2017. From 2017 onward, the social dynamics were activated and the simulation continued until 2200. We then computed the relative integrated absolute error (RIAE) between the two social model formulations for both the climate temperature anomaly $T(t)$ and the mean mitigation trajectory $\langle m(t) \rangle$, restricting the integral to the active period of social dynamics (2017–2200). Figure 5.2 shows the outcome of these sweeps, highlighting how the approximation error between binary and spectrum models depends on the underlying social parameters and the assumed critical temperature threshold. Overall, the results in Figure 5.2 demonstrate that the deviations between binary and spectrum replicator models remain small across a wide range of parameter values. For nearly all sweeps away from the baseline settings, the relative integrated absolute error (RIAE) of both the climate trajectory $T(t)$ and the mean mitigation trajectory $\langle m(t) \rangle$ stays below 15%. This indicates that

the binary formulation closely tracks the spectrum model even under substantial variation of social learning rate, reluctance, conformity, and maximum mitigation effort, confirming the robustness of the binary approximation in the coupled climate–social system. Full information on the simulation details, including the complete table of all parameter values used, is provided in Appendix D2.

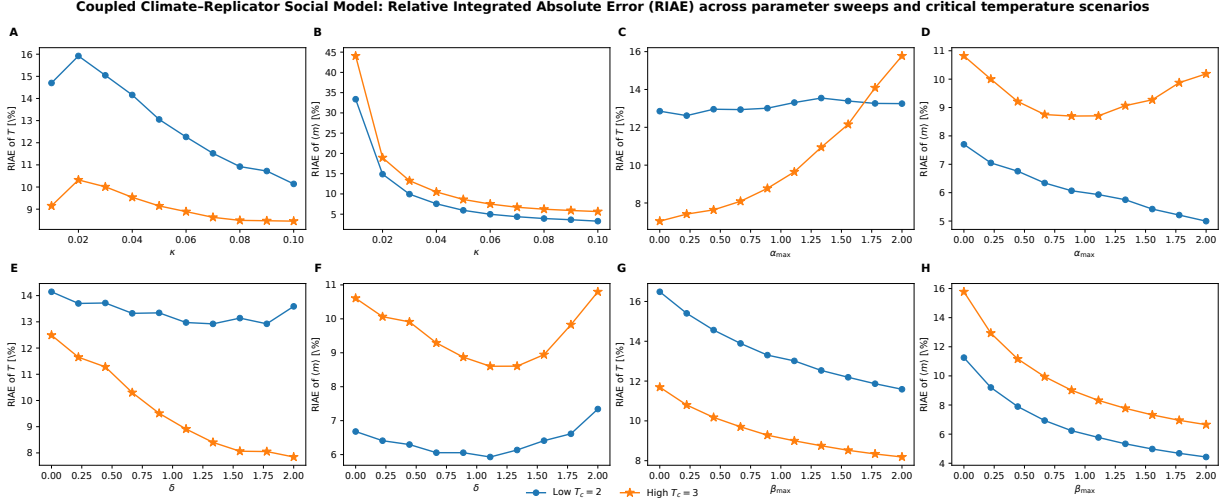


Figure 5.2: Illustration of binary ($N = 2$) and spectrum ($N = 100$) replicator models coupled to the climate–carbon system. Panels show RIAE of temperature anomaly $T(t)$ (A, C, E, and G) and mean mitigation $\langle m(t) \rangle$ (B, D, F, and H) across parameter sweeps for κ , $\alpha_{\max}$, δ , and $\beta_{\max}$.

5.4.2 Coupling Between the FJ Social Model and the Climate–Carbon Model

The coupling between the FJ social dynamics and the climate–carbon model is mediated by two feedbacks: temperature anomaly influences opinion change through a runaway forcing term, and the mean social opinion feeds back to alter effective emissions. In the FJ framework, the coupling to the climate system occurs through the runaway forcing term $R(T)$, defined as

$$R(T) = -R_0 + \frac{R_{\max}}{1 + \exp[-\alpha(T - T_c)]}, \quad (5.18)$$

where R_0 is a baseline offset, $R_{\max}$ is the saturation amplitude, α is the steepness, and T_c is a threshold temperature. As T increases beyond T_c , the logistic function grows, increasing

the pressure for opinions to shift toward pro-mitigation stances.

The collective outcome of the FJ dynamics is summarized by the average opinion. In the spectrum formulation this is:

$$\bar{X}_{\text{spec}}(t) = \frac{1}{n} \sum_{i=1}^n X_i(t), \quad (5.19)$$

whereas in the binary version only the sign of each opinion is retained,

$$\bar{X}_{\text{bin}}(t) = \frac{1}{n} \sum_{i=1}^n \text{sgn}(X_i(t)). \quad (5.20)$$

To systematically compare the binary ($N = 2$) and spectrum ($N = 100$) FJ social models when coupled to the same climate–carbon system, we carried out a set of controlled parameter sweeps with stochastic susceptibility and repeated runs per configuration. For each parameter setting, we generated paired trajectories under the continuous and binary formulations and compared the climate temperature anomaly $T(t)$ and the mean social signal $\langle m(t) \rangle$ and report the RIAE. We varied one parameter at a time over 10 evenly spaced values while holding the others fixed at baselines. The parameter ranges explored were:

- Mean susceptibility λ_0 : 0.2 to 0.8,
- Runaway amplitude R_{max} : 1.0 to 2.0,
- Adjustment rate Ψ : 0.01 to 0.10,
- Interaction sensitivity A : 0.2 to 1.0.

Baseline values when not swept were fixed at $R_0 = 0.778$, $R_{\text{max}} = 1.37$, $\Psi = 0.07$, $A = 0.35$, $\alpha = 5.7$, $\lambda_0 = 0.5$. Two scenarios were considered, with $T_c \in \{2.0, 3.0\}$ representing low and high collapse thresholds. These baseline values were inspired by the work in [112], which fitted the model parameters to empirical observations and identified the values that most accurately reproduce the recorded social–opinion dynamics. Individual susceptibilities λ_i are drawn at the start of each run from a normal distribution centered at the sweep mean λ_0 with variance 0.1, and values falling outside the interval $[0, 1]$ are clipped to the nearest boundary. The resulting heterogeneous λ_i values are then held fixed throughout the simulation. For each parameter value and each T_c scenario, we executed $N_{\text{reps}} = 10$ independent realizations (new draws of λ) for both the spectrum and binary formulations. We averaged the resulting RIAE values across the 10 runs and report the mean with error

bars equal to the standard error of the mean. Overall, the results depicted in Figure 5.3 shows the deviations between the binary and spectrum formulations remain small across the vast majority of sweep values; in nearly all cases the RIAE for both $T(t)$ and $\langle m(t) \rangle$ stays below 15%, indicating that the binary model closely tracks the spectrum model under wide variation of susceptibility, runaway amplitude, adjustment rate, and interaction sensitivity. An exception occurs at the high extreme of λ_0 , where the binary approximation begins to deviate more substantially from the spectrum formulation. Full information on simulation settings and the complete parameter table is provided in the Appendix D3.

Coupled Climate-FJ Social Model: Relative Integrated Absolute Error (RIAE) across parameter sweeps and T_c scenarios

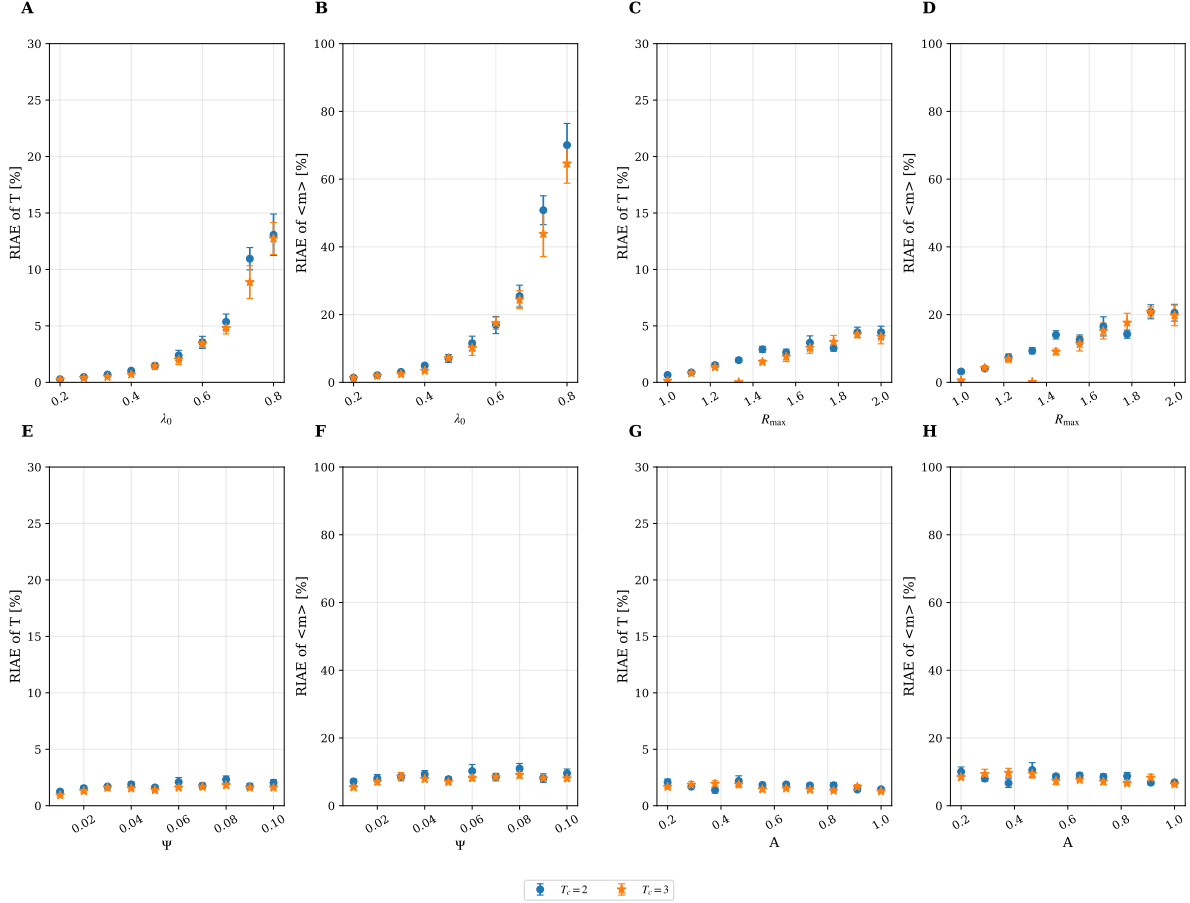


Figure 5.3: Illustration of binary and spectrum FJ social models coupled to the climate-carbon system. Panels show RIAE of $T(t)$ (A, C, E, and G) and $\langle m(t) \rangle$ (B, D, F, and H) across sweeps of mean susceptibility λ_0 , runaway amplitude $R_{\max}$, adjustment rate Ψ , and interaction sensitivity A .

5.4.3 Coupling of the Forest Model and the Replicator Social Model

In this section we combine the forest dynamics with the replicator social dynamics to form a coupled socio-ecological system. The coupling enters in two ways. First, individual

utilities in the social model depend on the ecological state:

$$U_i(F) = r_0 (2F - 1) \left(\frac{2i}{N-1} - 1 \right),$$

so that the payoffs associated with mitigation categories vary with the fraction of forest cover F . This formulation of utility, and its dependence on F , follows the approach introduced in [97]. Second, the replicator dynamics feed back into the ecological system through the mean social opinion $\langle m \rangle = \sum_i m_i x_i$, where $m_i \in [-1, 1]$ are fixed category opinions and x_i are their frequencies. This coupling enters the forest growth equation via a social feedback term

$$J(x) = h \langle m \rangle,$$

which shifts the balance of forest regeneration and loss depending on whether mitigation- or exploitation-oriented opinions dominate. To isolate the effect of social feedback, the simulations are run in two distinct phases. During the initial phase, the social model is switched off, allowing the forest fraction F to evolve autonomously until it reaches a quasi-equilibrium. At this point, the social dynamics are switched on, and the replicator system begins to co-evolve with the forest state. From this switching point onward, we measure RIAE between the binary ($N = 2$) and spectrum ($N = 100$) replicator formulations for both the forest fraction F and the mean opinion $\langle m \rangle$, continuing the simulation until the coupled system converges to a new equilibrium. Systematic parameter sweeps were performed across four key ecological and social parameters, while holding all others fixed at baseline values ($s = 0.1$, $r_0 = 1.0$, $v = 0.01$, $c = 0.3$, $k = 5.0$, $b = 0.0$). For the baseline configuration, we adopted parameter values reported in [97]. Specifically, we varied:

- the forest loss rate v from 0.0 to 0.12,
- the social learning rate κ from 0.01 to 0.10,
- the ecological feedback strength r_0 from 0.1 to 2.0,
- the forest regeneration parameter c from 0.1 to 0.5.

For each of these sweeps, we considered two values of the social–ecological coupling $h \in \{0.01, 0.1\}$, representing weak versus strong influence of mean opinion on forest growth. Figure 5.4 reports the results of these sweeps. Across most parameter ranges, the RIAE for both forest cover F and mean opinion $\langle m \rangle$ remains low, generally below 10%, demonstrating that the binary approximation captures the spectrum dynamics with high fidelity. Slightly higher deviations can occur under extreme parameter values (e.g. low r_0 or high

h), but overall the binary model tracks the spectrum formulation closely, even when forest and social dynamics are strongly coupled.

Coupled Forest-Replicator Social Model: Relative Integrated Absolute Error (RIAE) across parameter sweeps and h scenarios

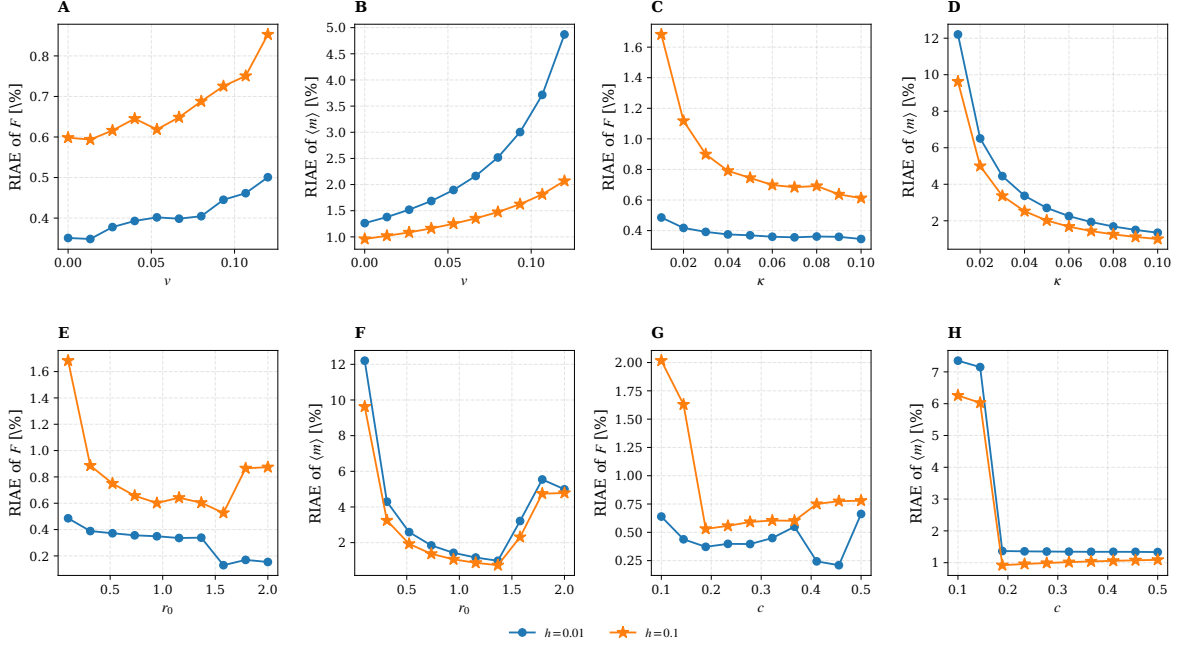


Figure 5.4: Illustration of binary and spectrum replicator models coupled to the forest–grassland system. RIAE of forest fraction F (A, C, E, and G) and mean opinion $\langle m \rangle$ (B, D, F, and H) are shown for sweeps of v , κ , r_0 , and c , under weak ($h = 0.01$) and strong ($h = 0.1$) coupling.

5.4.4 Coupling of the Forest Model and the FJ Social Model

The coupling between the forest dynamics and the FJ social dynamics is bidirectional and occurs through two feedback channels. First, the forest state F influences social dynamics through the runaway feedback term,

$$R(T) = -R_0 + \frac{R_{\max}}{1 + \exp[-\alpha(F - F_c)]}, \quad (5.21)$$

which enters additively into the FJ opinion update. This term biases individuals toward more pro-ecological stances as the forest fraction approaches or falls below the critical threshold F_c .

Second, the mean social opinion feeds back into the forest through

$$J(x) = h \bar{X}(t), \quad \bar{X}(t) = \begin{cases} \frac{1}{N} \sum_{i=1}^N X_i(t), & \text{continuous case} \\ \frac{1}{N} \sum_{i=1}^N \text{sign}(X_i(t)), & \text{binary case} \end{cases} \quad (5.22)$$

which modifies the forest growth equation by adding or subtracting from the net regeneration rate depending on whether the average opinion supports or resists conservation. From this switching point onward, we compute the relative integrated absolute error (RIAE) between the binary ($N = 2$) and spectrum ($N = 100$) FJ formulations for both the forest fraction F and the mean opinion $\langle m \rangle$, continuing the simulation until the coupled system converges to a new equilibrium.

Unless being swept, parameters are fixed at

$$R_0 = 1, \alpha = 5, k = 5, h = 0.01, R_{\max} = 1.5, \Psi = 0.05, A = 0.35, \lambda_0 = 0.5, c = 0.3, v = 0.01,$$

We examine two ecological threshold scenarios

$$F_c \in \{0.2, 0.8\},$$

representing a resilient versus fragile forest regime. We then vary, one at a time:

- $R_{\max}$: 1.0 to 2.0,
- Ψ : 0.01 to 0.10,
- A : 0.2 to 1.0,
- λ_0 : 0.2 to 0.8,
- c : 0.1 to 0.5,
- v : 0.0 to 0.12.

For each grid point we run 10 repeated realizations and report the mean RIAE with standard-error bars. Figure 5.5 shows the outcomes of the parameter sweeps for the coupled Forest–FJ system. In nearly all cases, the deviation between the binary and spectrum models for the forest fraction F remains minimal when the parameters lie within intermediate ranges. For the social layer, the relative differences in mean opinion $\langle X \rangle$ are generally higher than for the forest dynamics, yet still remain within a range that provides a reliable approximation. Overall, these results demonstrate that the binary FJ formulation

can capture the essential coupled dynamics with good fidelity across a wide spectrum of parameter regimes.

Coupled Forest-FJ Social Model: Relative Integrated Absolute Error (RIAE) across parameter sweeps and F_c scenarios

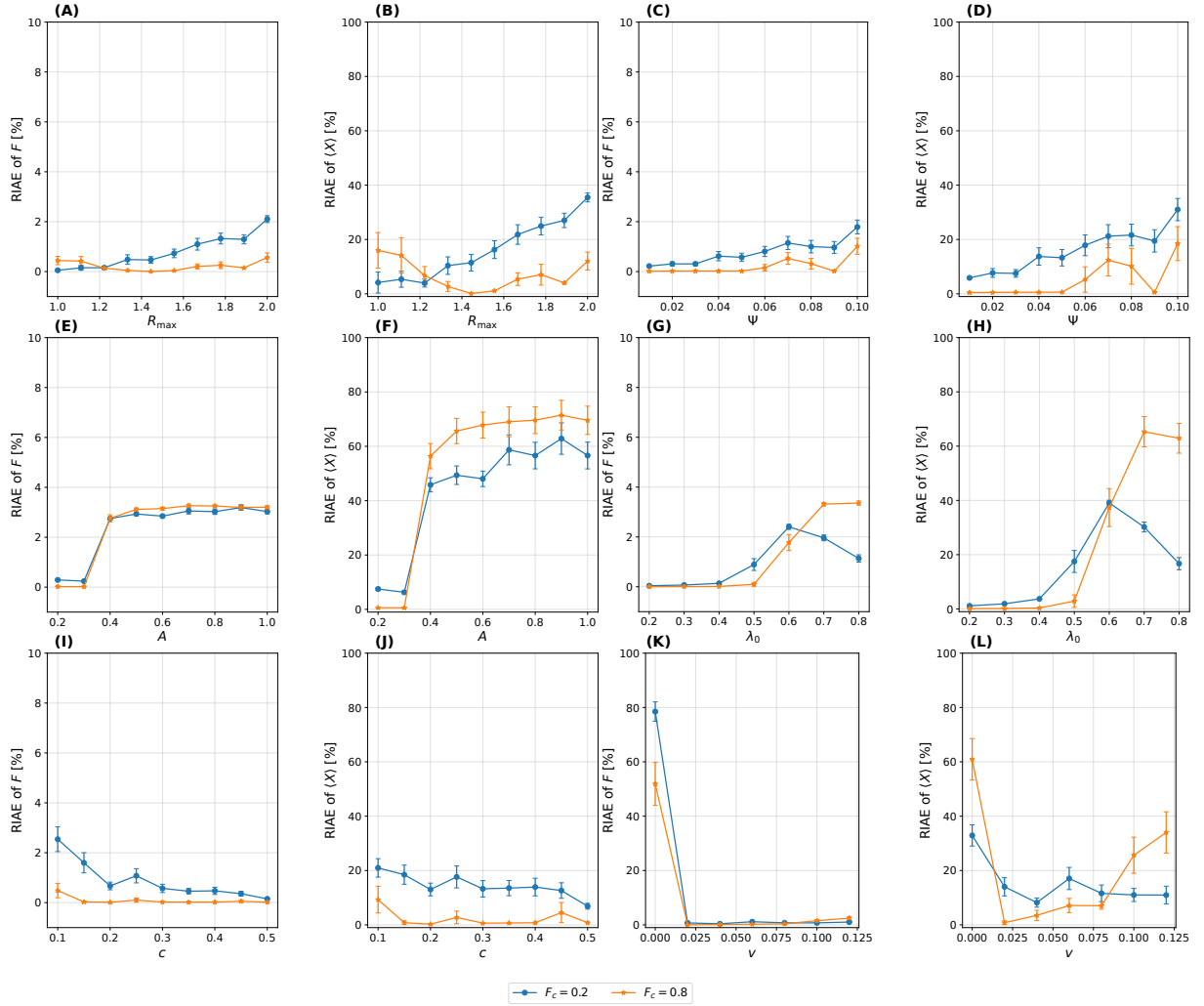


Figure 5.5: Illustration of binary and spectrum FJ models coupled to the forest–grassland system. Panels report RIAE of forest fraction F (A, C, E, G, I, and K) and mean opinion $\langle X \rangle$ (B, D, F, H, J, and L) across sweeps of $R_{\max}$, Ψ , A , λ_0 , c , and v , under two ecological thresholds $F_c \in \{0.2, 0.8\}$.

5.5 Discussion

In this work, our aim is not to claim that binary social models replicate the full behaviour of continuous formulations. Instead, we quantify how well the binary model approximates the behaviour of the full spectrum model across parameter regimes. Across the four couplings we examined, replicator vs. Friedkin–Johnsen dynamics paired with either the climate–carbon system or the forest–grassland mosaic, we identify broad regions where the binary model tracks its spectrum counterpart with errors typically within $\sim 15\%$ RIAE, and often lower.

This result should be read as qualified robustness, not equivalence. Continuous models retain descriptive richness and can deviate from binary behaviour outside the identified ranges or near sensitive thresholds. Nevertheless, by quantifying where the approximation holds, and its typical magnitude (RIAE $\lesssim 15\%$ in most cases), we provide a practical guide for when the computational simplicity and interpretability of binary models can be leveraged without materially distorting system-level conclusions.

Despite these encouraging results, several limitations should be noted. First, our models necessarily involve simplifications of both ecological and social processes. For instance, social dynamics are assumed to be homogeneous within categories and to follow fixed interaction rules, whereas in reality, social influence can be heterogeneous, adaptive, and context-dependent. Likewise, ecological models are stylized representations of climate and forest systems, omitting spatial heterogeneity, stochastic disturbances, and multi-species interactions. Second, our evaluation focuses on long-term equilibrium behavior and relative trajectory agreement. More transient phenomena, such as short-term oscillations, abrupt cascades, or early-warning signals, may reveal subtler differences between binary and spectrum formulations. Finally, the RIAE metric, though well-suited to capturing integrated deviations, does not directly assess potential differences in variance, distributional properties, or extreme outcomes.

Future work can address these limitations by incorporating richer forms of heterogeneity and complexity into both the social and ecological layers. For example, integrating network-based social structures, adaptive susceptibility, or multi-dimensional opinions could reveal conditions under which spectrum models diverge more strongly from binary approximations. On the ecological side, extending to spatially explicit landscapes or Earth system models of higher complexity would test whether binary models remain robust at larger scales. In addition, future studies could examine the performance of binary models in capturing leading indicators of critical transitions, such as flickering or rising variance, which are not the focus of the present work. Finally, empirical validation through comparison with observed social and ecological time series would help assess the practical utility of binary models for forecasting and policy design.

Chapter 6

Conclusion

This thesis developed a unified framework for detecting, characterising, and modelling critical transitions in socio-climate and complex systems by integrating dynamical systems theory, stochastic analysis, and deep learning. Across four chapters, we combined mathematical models, computational methods, and empirical validation to advance the understanding of how tipping points emerge, how early-warning signals may be detected, and how human social behaviour interacts with climate thresholds.

In Chapter 2, we introduced a deep learning approach for bifurcation detection using a hybrid CNN-LSTM architecture trained on synthetic time series derived from canonical normal-form systems. The model successfully learned universal dynamical features of transcritical, fold, and Hopf bifurcations, achieving high accuracy and outperforming traditional early-warning indicators. Importantly, it generalised to empirical data including palaeoclimate and laboratory datasets, demonstrating that data-driven methods can reliably classify approaching transitions even under coloured noise.

Chapter 3 extended these ideas to the detection of flickering, a stochastic early-warning signal arising from noise-induced switching between coexisting states. We constructed a large-scale simulation dataset spanning multiple bistable dynamical systems and trained a deep learning ensemble that distinguished true flickering from noise-modulated null dynamics. The classifier consistently outperformed variance-based indicators across both simulated and empirical datasets, including hibernating-animal temperature series and palaeoclimate records. These results show that deep learning can uncover subtle, multi-scale patterns of declining resilience that conventional statistics often miss.

In Chapter 4, we developed a coupled socio-climate model that links behavioural dynamics, emissions, carbon-cycle feedbacks, and global temperature. We showed that social processes, such as learning rate, perceived mitigation cost, and normative reinforcement, can delay, accelerate, or prevent climate tipping points. Faster social learning and stronger

norms promoted stabilising “social tipping,” whereas sluggish behavioural adjustment increased vulnerability to climate thresholds. Across scenarios, the interaction between social responsiveness and climatic feedbacks produced a rich landscape of possible futures, highlighting that social processes can be as influential as physical ones in determining whether climate transitions occur.

Chapter 5 evaluated whether binary opinion models can adequately approximate spectrum models in coupled human–environment systems. Using replicator and Friedkin–Johnsen frameworks connected to climate–carbon and forest–grassland models, we quantified the discrepancy between binary and spectrum formulations through the Relative Integrated Absolute Error (RIAE). Results showed that binary models typically reproduce system-level behaviour with errors below 15% in most parameter regimes, particularly under moderate susceptibility, intermediate ecological turnover, and away from runaway forcing. Despite their reduced dimensionality, binary models captured key feedbacks, tipping behaviour, and collective dynamics with high fidelity, suggesting that they are often sufficient for large-scale socio–ecological modelling.

Taken together, the results of this thesis demonstrate that combining deep learning with dynamical-systems models provides a powerful methodology for anticipating and understanding critical transitions. Methodologically, we showed that machine learning can detect early-warning signals—including bifurcations and flickering, with robustness to noise structure, limited data, and model uncertainty. Conceptually, we established that social dynamics play a decisive role in determining whether climate tipping points are reached or avoided, and that simplified behavioural models can often approximate more complex opinion distributions without losing critical predictive power.

Future work may extend these approaches by incorporating higher-dimensional climate models, heterogeneous social networks, adaptive policy interventions, and real-time data assimilation. Deep learning frameworks may also be expanded to perform sequence-to-sequence prediction of tipping events or uncertainty quantification. Ultimately, the integration of theory-driven and data-driven tools offers a promising path toward improved monitoring, prediction, and mitigation of critical transitions in coupled human–environment systems.

References

- [1] Daron Acemoglu and Asuman Ozdaglar. Opinion dynamics and learning in social networks. Dynamic Games and Applications, 1:3–49, 2011.
- [2] Linda JS Allen. An introduction to stochastic processes with applications to biology. CRC press, 2010.
- [3] Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. Journal of big Data, 8:1–74, 2021.
- [4] Brian DO Anderson and Mengbin Ye. Recent advances in the modelling and analysis of opinion dynamics on influence networks. International Journal of Automation and Computing, 16(2):129–149, 2019.
- [5] Philip W Anderson. More is different: broken symmetry and the nature of the hierarchical structure of science. Science, 177(4047):393–396, 1972.
- [6] Koon Meng Ang, El-Sayed M El-Kenawy, Abdelaziz A Abdelhamid, Abdelhameed Ibrahim, Amal H Alharbi, Doaa Sami Khafaga, Sew Sun Tiang, and Wei Hong Lim. Optimal design of convolutional neural network architectures using teaching-learning-based optimization for image classification. Symmetry, 14(11):2323, 2022.
- [7] David I Armstrong McKay, Arie Staal, Jesse F Abrams, Ricarda Winkelmann, Boris Sakschewski, Sina Loriani, Ingo Fetzer, Sarah E Cornell, Johan Rockström, and Timothy M Lenton. Exceeding 1.5 c global warming could trigger multiple climate tipping points. Science, 377(6611):eabn7950, 2022.
- [8] Vladimir Igorevich Arnold. Geometrical methods in the theory of ordinary differential equations, volume 250. Springer Science & Business Media, 2012.

- [9] Peter Ashwin, Sebastian Wieczorek, Renato Vitolo, and Peter Cox. Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370(1962):1166–1184, 2012.
- [10] Yazdan Babazadeh Maghsoodlo, Madhur Anand, and Chris T Bauch. Social dynamics can delay or prevent climate tipping points by speeding the adoption of climate change mitigation. Proceedings of the Royal Society A, 481(2322):20250376, 2025.
- [11] Yazdan Babazadeh Maghsoodlo, Daniel Dylewsky, Madhur Anand, and Chris T. Bauch. Deep learning for bifurcation detection: Extending early warning signals to dynamical systems with coloured noise. Mathematics, 13(17), 2025.
- [12] Yazdan Babazadeh Maghsoodlo, Amin Safaeesirat, and Fakhteh Ghanbarnejad. The big bang of an epidemic: a metapopulation approach to identify the spatiotemporal origin of contagious diseases and their universal spreading pattern. Scientific Reports, 15(1):5809, 2025.
- [13] Lee-Ann Barlow, Jacob Cecile, Chris T Bauch, and Madhur Anand. Modelling interactions between forest pest invasions and human decisions regarding firewood transport restrictions. PloS one, 9(4):e90511, 2014.
- [14] Igor A Bashmakov, Lars J Nilsson, Adolf Acquaye, Christopher Bataille, Jonathan M Cullen, Manfred Fischedick, Yong Geng, Kanako Tanaka, et al. Climate change 2022: Mitigation of climate change. contribution of working group iii to the sixth assessment report of the intergovernmental panel on climate change, chapter 11. 2022.
- [15] Chris T Bauch. Imitation dynamics predict vaccinating behaviour. Proceedings of the Royal Society B: Biological Sciences, 272(1573):1669–1675, 2005.
- [16] Max Bechthold, Wolfram Barfuss, André Butz, Jannes Breier, Sara M Constantino, Jobst Heitzig, Luana Schwarz, Sanam N Vardag, and Jonathan F Donges. Social norms and groups structure safe operating spaces in renewable resource use in a social-ecological multi-layer network model. EGUsphere, 2024:1–43, 2024.
- [17] Brian Beckage, Louis J Gross, Katherine Lacasse, Eric Carr, Sara S Metcalf, Jonathan M Winter, Peter D Howe, Nina Fefferman, Travis Franck, Asim Zia, et al. Linking models of human behaviour and climate alters projected climate change. Nature Climate Change, 8(1):79–84, 2018.

- [18] Brian Beckage, Katherine Lacasse, Jonathan M Winter, Louis J Gross, Nina Feferman, Forrest M Hoffman, Sara S Metcalf, Travis Franck, Eric Carr, Asim Zia, et al. The earth has humans, so why don't our climate models? Climatic Change, 163:181–188, 2020.
- [19] Brian Beckage, Frances C Moore, and Katherine Lacasse. Incorporating human behaviour into earth system modelling. Nature Human Behaviour, 6(11):1493–1502, 2022.
- [20] Roberto Benzi, Alfonso Sutera, and Angelo Vulpiani. The mechanism of stochastic resonance. Journal of Physics A: mathematical and general, 14(11):L453, 1981.
- [21] Carmela Bernardo, Lingfei Wang, Francesco Vasca, Yiguang Hong, Guodong Shi, and Claudio Altafini. Achieving consensus in multilateral international negotiations: The case study of the 2015 paris agreement on climate change. Science Advances, 7(51):eabg8068, 2021.
- [22] Claudia R Binder, Jochen Hinkel, Pieter WG Bots, and Claudia Pahl-Wostl. Comparison of frameworks for analyzing social-ecological systems. Ecology and society, 18(4), 2013.
- [23] Elisabetta Biondi, Chiara Boldrini, Andrea Passarella, and Marco Conti. Dynamics of opinion polarization. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 53(9):5381–5392, 2023.
- [24] Soumyajyoti Biswas. Mean-field solutions of kinetic-exchange opinion models. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 84(5):056106, 2011.
- [25] Carl Boettiger and Alan Hastings. Early warning signals and the prosecutor's fallacy. Proceedings of the Royal Society B: Biological Sciences, 279(1748):4734–4739, 2012.
- [26] Carl Boettiger and Alan Hastings. Quantifying limits to detection of early warning for critical transitions. Journal of the Royal Society Interface, 9(75):2527–2539, 2012.
- [27] Lovely Borah, Kaushik Dehingia, Hemanta Kumar Sarmah, Animesh Phukan, Fathalla Ali Rihan, and Evren Hınçal. Stability and bifurcation analysis of a financial dynamical system with time delay. The European Physical Journal Special Topics, pages 1–16, 2025.

- [28] Kaitlyn Boudreau, Melanie Robinson, and Zahrah Farooqi. Ipcc sixth assessment report: Climate change 2021: The physical science basis summary for policymakers. Canadian Journal of Emergency Management, 2(1), 2022.
- [29] Alberto Bressan. Tutorial on the center manifold theorem. Hyperbolic systems of balance laws, 1911:327–344, 2003.
- [30] Thomas M Bury, Chris T Bauch, and Madhur Anand. Charting pathways to climate change mitigation in a coupled socio-climate model. PLoS computational biology, 15(6):e1007000, 2019.
- [31] Thomas M. Bury, Chris T. Bauch, and Madhur Anand. Charting pathways to climate change mitigation in a coupled socio-climate model. PLOS Computational Biology, 15(6):1–16, 06 2019.
- [32] Thomas M Bury, Chris T Bauch, and Madhur Anand. Detecting and distinguishing tipping points using spectral early warning signals. Journal of the Royal Society Interface, 17(170):20200482, 2020.
- [33] Thomas M Bury, Daniel Dylewsky, Chris T Bauch, Madhur Anand, Leon Glass, Alvin Shrier, and Gil Bub. Predicting discrete-time bifurcations with deep learning. Nature Communications, 14(1):6331, 2023.
- [34] Thomas M Bury, RI Sujith, Induja Pavithran, Marten Scheffer, Timothy M Lenton, Madhur Anand, and Chris T Bauch. Deep learning for early warning signals of tipping points. Proceedings of the National Academy of Sciences, 118(39):e2106140118, 2021.
- [35] Stephen R Carpenter and William A Brock. Rising variance: a leading indicator of ecological transition. Ecology letters, 9(3):311–318, 2006.
- [36] Jack Carr. Applications of centre manifold theory, volume 35. Springer Science & Business Media, 2012.
- [37] Damon Centola, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. Experimental evidence for tipping points in social convention. Science, 360(6393):1116–1119, 2018.
- [38] Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. Analog forecasting of extreme-causing weather patterns using deep learning. Journal of Advances in Modeling Earth Systems, 12(2):e2019MS001958, 2020.

- [39] Hamish Clarke, Rachael H Nolan, Victor Resco De Dios, Ross Bradstock, Anne Griebel, Shiva Khanal, and Matthias M Boer. Forest fire threatens global carbon sinks and population centres under rising atmospheric water demand. Nature communications, 13(1):7161, 2022.
- [40] John David Crawford. Introduction to bifurcation theory. Reviews of modern physics, 63(4):991, 1991.
- [41] Vasilis Dakos, Chris A Boulton, Josh E Buxton, Jesse F Abrams, David I Armstrong McKay, Sebastian Bathiany, Lana Blaschke, Niklas Boers, Daniel Dylewsky, Carlos López-Martínez, et al. Tipping point detection and early-warnings in climate, ecological, and human systems. EGUsphere, 2023:1–35, 2023.
- [42] Vasilis Dakos, Stephen R Carpenter, William A Brock, Aaron M Ellison, Vishwesha Guttal, Anthony R Ives, Sonia Kéfi, Valerie Livina, David A Seekell, Egbert H van Nes, et al. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. PloS one, 7(7):e41010, 2012.
- [43] Vasilis Dakos, Marten Scheffer, Egbert H Van Nes, Victor Brovkin, Vladimir Petoukhov, and Hermann Held. Slowing down as an early warning signal for abrupt climate change. Proceedings of the National Academy of Sciences, 105(38):14308–14312, 2008.
- [44] Vasilis Dakos, Egbert H Van Nes, Paolo d’Odorico, and Marten Scheffer. Robustness of variance and autocorrelation as indicators of critical slowing down. Ecology, 93(2):264–271, 2012.
- [45] Vasilis Dakos, Egbert H van Nes, and Marten Scheffer. Flickering as an early warning signal. Theoretical ecology, 6:309–317, 2013.
- [46] Smita Deb, Sahil Sidheekh, Christopher F Clements, Narayanan C Krishnan, and Partha S Dutta. Machine learning methods trained on simple models can predict critical transitions in complex natural systems. Royal Society Open Science, 9(2):211475, 2022.
- [47] Robert M DeConto and David Pollard. Contribution of antarctica to past and future sea-level rise. Nature, 531(7596):591–597, 2016.
- [48] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. Advances in Complex Systems, 3(01n04):87–98, 2000.

- [49] Morris H DeGroot. Reaching a consensus. Journal of the American Statistical association, 69(345):118–121, 1974.
- [50] Mark M Dekker, Anna S Von Der Heydt, and Henk A Dijkstra. Cascading transitions in the climate system. Earth System Dynamics, 9(4):1243–1260, 2018.
- [51] O Descalzi, S Curilef, L Velazquez, and V Muñoz. Complex systems and inter/trans-disciplinary research: A review. Chaos: An Interdisciplinary Journal of Nonlinear Science, 34(1), 2024.
- [52] Peter D Ditlevsen. Observation of α -stable noise induced millennial climate changes from an ice-core record. Geophysical Research Letters, 26(10):1441–1444, 1999.
- [53] Peter D Ditlevsen and Sigfus J Johnsen. Tipping points: early warning and wishful thinking. Geophysical Research Letters, 37(19), 2010.
- [54] Ian Dobson, Benjamin A Carreras, Vickie E Lynch, and David E Newman. Complex systems analysis of series of blackouts: Cascading failure, critical points, and self-organization. Chaos: An Interdisciplinary Journal of Nonlinear Science, 17(2), 2007.
- [55] Scott C Doney, Victoria J Fabry, Richard A Feely, and Joan A Kleypas. Ocean acidification: the other co2 problem. Annual review of marine science, 1(1):169–192, 2009.
- [56] John M Drake and Blaine D Griffen. Early warning signals of extinction in deteriorating environments. Nature, 467(7314):456–459, 2010.
- [57] Daniel Dylewsky, Madhur Anand, and Chris T Bauch. Early warning signals for bifurcations embedded in high dimensions. Scientific Reports, 14(1):18277, 2024.
- [58] Daniel Dylewsky, Sonia Kéfi, Madhur Anand, and Chris T Bauch. Neural models for prediction of spatially patterned phase transitions: methods and challenges. Theoretical Ecology, 18(1):1–13, 2025.
- [59] Daniel Dylewsky, Timothy M Lenton, Marten Scheffer, Thomas M Bury, Christopher G Fletcher, Madhur Anand, and Chris T Bauch. Universal early warning signals of phase transitions in climate systems. Journal of the Royal Society Interface, 20(201):20220562, 2023.
- [60] Ullrich Ecker, Matthew Hornsey, and Stephan Lewandowsky. Rapid shifts in public support for climate action following extreme events. Nature Climate Change, 15(2):312–320, 2025.

- [61] Sibel Eker, Timothy M Lenton, Tom Powell, Jürgen Scheffran, Steven R Smith, Deepthi Swamy, and Caroline Zimm. Cross-system interactions for positive tipping cascades. Earth System Dynamics, 15(3):789–800, 2024.
- [62] Jeffrey L Elman. Finding structure in time. Cognitive science, 14(2):179–211, 1990.
- [63] Hans Engler, Hans G Kaper, Tasso J Kaper, and Theodore Vo. Modeling the dynamics of glacial cycles. Mathematics of Planet Earth: Protecting Our Planet, Learning from the Past, Safeguarding for the Future, pages 3–33, 2019.
- [64] Ernesto Estrada. What is a complex system, after all? Foundations of Science, 29(4):1143–1170, 2024.
- [65] Nikolaos Evangelou, Tianqi Cui, Juan M Bello-Rivas, Alexei Makeev, and Ioannis G Kevrekidis. Tipping points of evolving epidemiological networks: Machine learning-assisted, data-driven effective modeling. Chaos: An Interdisciplinary Journal of Nonlinear Science, 34(6), 2024.
- [66] Isaiah Farahbakhsh, Chris T Bauch, and Madhur Anand. Modelling coupled human–environment complexity for the future of the biosphere: strengths, gaps and promising directions. Philosophical Transactions of the Royal Society B, 377(1857):20210382, 2022.
- [67] Isaiah Farahbakhsh, Chris T Bauch, and Madhur Anand. Drivers of tipping points in coupled human-environment systems. EGUsphere, 2023:1–23, 2023.
- [68] Isaiah Farahbakhsh, Chris T Bauch, and Madhur Anand. Tipping points in coupled human–environment system models: a review. Earth System Dynamics, 15(4):947–967, 2024.
- [69] Samer Fawzy, Ahmed I Osman, John Doran, and David W Rooney. Strategies for mitigation of climate change: a review. Environmental Chemistry Letters, 18:2069–2094, 2020.
- [70] Thomas Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. European journal of operational research, 270(2):654–669, 2018.
- [71] Carl Folke, Åsa Jansson, Johan Rockström, Per Olsson, Stephen R Carpenter, F Stuart Chapin III, Anne-Sophie Crépin, Gretchen Daily, Kjell Danell, Jonas Ebbesson, et al. Reconnecting to the biosphere. Ambio, 40(7):719–738, 2011.

- [72] I Freidlin Mark, Szücs Joseph, and D Wentzell Alexander. Random perturbations of dynamical systems. Grundlehren math. Wiss, 260, 2012.
- [73] Tak Fung, James P O’Dwyer, and Ryan A Chisholm. Species-abundance distributions under colored environmental noise. Journal of mathematical biology, 74:289–311, 2017.
- [74] Dennis Gilfillan and Gregg Marland. Cdiac-ff: global and national co 2 emissions from fossil fuel combustion and cement manufacture: 1751–2017. Earth System Science Data, 13(4):1667–1680, 2021.
- [75] David Gilljam, Jonas Knape, Andreas Lindén, Marianne Mugabo, Steven M Sait, and Mike S Fowler. The colour of environmental fluctuations associated with terrestrial animal population dynamics. Global Ecology and Biogeography, 28(2):118–130, 2019.
- [76] Matt Grove, Lucy Timbrell, Ben Jolley, Fiona Polack, and James M Borg. The importance of noise colour in simulations of evolutionary systems. Artificial Life, 27(3–4):164–182, 2021.
- [77] John Guckenheimer and Philip Holmes. Nonlinear oscillations, dynamical systems, and bifurcations of vector fields, volume 42. Springer Science & Business Media, 2013.
- [78] Vishwesha Guttal and Ciriya Jayaprakash. Changing skewness: an early warning signal of regime shifts in ecosystems. Ecology letters, 11(5):450–460, 2008.
- [79] Klaus Hasselmann. Stochastic climate models part i. theory. tellus, 28(6):473–485, 1976.
- [80] Zitao He and Chris T Bauch. Effect of homophily on coupled behavior-disease dynamics near a tipping point. Mathematical Biosciences, 376:109264, 2024.
- [81] James B Heaton, Nick G Polson, and Jan Hendrik Witte. Deep learning for finance: deep portfolios. Applied Stochastic Models in Business and Industry, 33(1):3–12, 2017.
- [82] C Heinze, V Eyring, P Friedlingstein, C Jones, Y Balkanski, W Collins, T Fichefet, S Gao, A Hall, D Ivanova, et al. Esd reviews: Climate feedbacks in the earth system and prospects for their evaluation, earth syst. dynam., 10, 379–452, 2019.

- [83] Kirsten A Henderson, Chris T Bauch, and Madhur Anand. Alternative stable states and the sustainability of forests, grasslands, and agriculture. Proceedings of the National Academy of Sciences, 113(51):14552–14559, 2016.
- [84] Rick Hennekam, Bregje Van der Bolt, Egbert van Nes, Gert J de Lange, Marten Scheffer, and Gert Jan Reichart. Calibrated xrf-scanning data (mm resolution) and calibration data (icp-oes and icp-ms) for elements al, ba, mo, ti, and u in mediterranean cores ms21, ms66, and 64pe406e1. 2020.
- [85] Rick Hennekam, Bregje van der Bolt, Egbert H van Nes, Gert J de Lange, Marten Scheffer, and Gert-Jan Reichart. Early-warning signals for marine anoxic events. Geophysical Research Letters, 47(20):e2020GL089183, 2020.
- [86] Marina Hirota, Milena Holmgren, Egbert H Van Nes, and Marten Scheffer. Global resilience of tropical forest and savanna to critical transitions. Science, 334(6053):232–235, 2011.
- [87] Morris W Hirsch, Stephen Smale, and Robert L Devaney. Differential equations, dynamical systems, and an introduction to chaos. Academic press, 2013.
- [88] S Hochreiter. Long short-term memory. Neural Computation MIT-Press, 1997.
- [89] David Holcman and Misha Tsodyks. The emergence of up and down states in cortical networks. PLoS computational biology, 2(3):e23, 2006.
- [90] John H Holland. Emergence: From chaos to order. OUP Oxford, 2000.
- [91] Mehran Homayounfar and Rachata Muneeppeerakul. On coupled dynamics and regime shifts in coupled human–water systems. Hydrological Sciences Journal, 66(5):769–776, 2021.
- [92] Werner Horsthemke. Noise induced transitions. In Stochastic Nonlinear Systems in Physics, Chemistry, and Biology: Proceedings of the Workshop Bielefeld, Fed. Rep. of Germany, October 5–11, 1980, pages 116–126. Springer, 1985.
- [93] Solomon M Hsiang, Marshall Burke, and Edward Miguel. Quantifying the influence of climate on human conflict. Science, 341(6151):1235367, 2013.
- [94] Yu Huang, Sebastian Bathiany, Peter Ashwin, and Niklas Boers. Deep learning for predicting rate-induced tipping. Nature Machine Intelligence, pages 1–10, 2024.

- [95] Terence P Hughes. Catastrophes, phase shifts, and large-scale degradation of a caribbean coral reef. Science, 265(5178):1547–1551, 1994.
- [96] Syed Nazir Hussain, Azlan Abd Aziz, Md Jakir Hossen, Nor Azlina Ab Aziz, G Ramana Murthy, and Fajaruddin Bin Mustakim. A novel framework based on cnn-lstm neural network for prediction of missing values in electricity consumption time-series datasets. Journal of Information Processing Systems, 18(1):115–129, 2022.
- [97] Clinton Innes, Madhur Anand, and Chris T Bauch. The impact of human-environment interactions on the stability of forest-grassland mosaic ecosystems. Scientific reports, 3(1):2689, 2013.
- [98] International Renewable Energy Agency (IRENA). Accelerating renewable energy deployment: The role of the g20 in driving global change. Technical report, International Renewable Energy Agency, 2019. Accessed: 2025-11-07.
- [99] David Jacobs and Robert Atkinson. The decline and transformation of mature industries: Lessons from industrial evolution. Business History, 60(5):742–761, 2017.
- [100] Peter C Jentsch, Chris T Bauch, and Madhur Anand. Fire mitigates bark beetle outbreaks in serotinous forests. Theoretical Ecology, 14(4):611–621, 2021.
- [101] Daniel Jercog, Alex Roxin, Peter Bartho, Artur Luczak, Albert Compte, and Jaime De La Rocha. Up-down cortical dynamics reflect state transitions in a bistable network. Elife, 6:e22425, 2017.
- [102] Louis Jezequel and Claude-Henri Lamarque. Analysis of non-linear dynamical systems by the normal form theory. Journal of sound and vibration, 149(3):429–459, 1991.
- [103] Matthew W Jones, Douglas I Kelley, Chantelle A Burton, Francesca Di Giuseppe, Maria Lucia F Barbosa, Esther Brambleby, Andrew J Hartley, Anna Lombardi, Guilherme Mataveli, Joe R McNorton, et al. State of wildfires 2023–2024. Earth System Science Data, 16(8):3601–3685, 2024.
- [104] Marko Jusup, Petter Holme, Kiyoshi Kanazawa, Misako Takayasu, Ivan Romić, Zhen Wang, Sunčana Geček, Tomislav Lipić, Boris Podobnik, Lin Wang, et al. Social physics. Physics Reports, 948:1–148, 2022.
- [105] Ph.D. Jyoti Dabass. Supercharge your nlp apps with xlstm: Enhancing accuracy capacity, May 2024. Accessed: 2025-11-07.

- [106] Rebecca Killick, Paul Fearnhead, and Idris A Eckley. Optimal detection of change-points with a linear computational cost. Journal of the American Statistical Association, 107(500):1590–1598, 2012.
- [107] Yun-Long Kong, Qingqing Huang, Chengyi Wang, Jingbo Chen, Jiansheng Chen, and Dongxu He. Long short-term memory neural networks for online disturbance detection in satellite image time series. Remote Sensing, 10(3):452, 2018.
- [108] Marz Kozłowska, Mateusz Denys, Mateusz Wiliński, Grzegorz Link, Tomasz Gubiec, Tomasz R Werner, Ryszard Kutner, and Zbigniew R Struzik. Dynamic bifurcations on financial markets. Chaos, Solitons & Fractals, 88:126–142, 2016.
- [109] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 2012.
- [110] Christian Kuehn. A mathematical framework for critical transitions: Bifurcations, fast-slow systems and stochastic dynamics. Physica D: Nonlinear Phenomena, 240(12):1020–1035, 2011.
- [111] Christian Kuehn, Gerd Zschaler, and Thilo Gross. Early warning signs for saddle-escape transitions in complex networks. Scientific reports, 5(1):13190, 2015.
- [112] Athira Satheesh Kumar, Krešimir Josić, Chris T Bauch, and Madhur Anand. From opinion polarization to climate action: A social-climate model of the opinion spectrum. arXiv preprint arXiv:2503.04689, 2025.
- [113] Yuri A Kuznetsov. Elements of applied bifurcation theory. Springer, 1998.
- [114] Yuri A Kuznetsov. Saddle-node bifurcation. Scholarpedia, 1(10):1859, 2006.
- [115] CAK Kwuimy, Foad Nazari, Xun Jiao, Pejman Rohani, and C Nataraj. Nonlinear dynamic analysis of an epidemiological model for covid-19 including public behavior and government action. Nonlinear Dynamics, 101:1545–1559, 2020.
- [116] Steven J Lade and Thilo Gross. Early warning signals for critical transitions: a generalized modeling approach. PLoS computational biology, 8(2):e1002360, 2012.
- [117] Steven J Lade, Alessandro Tavoni, Simon A Levin, and Maja Schlüter. Regime shifts in a social-ecological system. Theoretical ecology, 6(3):359–372, 2013.

- [118] Russell Lande, Steinar Engen, and Bernt-Erik Saether. Stochastic population dynamics in ecology and conservation. OUP Oxford, 2003.
- [119] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. nature, 521(7553):436–444, 2015.
- [120] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 2002.
- [121] Stephen Legg. Ipcc, 2021: Climate change 2021-the physical science basis. Interaction, 49(4):44–45, 2021.
- [122] Timothy M Lenton. Land and ocean carbon cycle feedback effects on global warming in a simple earth system model. Tellus B: Chemical and Physical Meteorology, 52(5):1159–1188, 2000.
- [123] Timothy M Lenton. Arctic climate tipping points. Ambio, 41(1):10–22, 2012.
- [124] Timothy M Lenton, Hermann Held, Elmar Kriegler, Jim W Hall, Wolfgang Lucht, Stefan Rahmstorf, and Hans Joachim Schellnhuber. Tipping elements in the earth’s climate system. Proceedings of the national Academy of Sciences, 105(6):1786–1793, 2008.
- [125] Timothy M Lenton, VN Livina, Vasilis Dakos, Egbert H van Nes, and M Scheffer. Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370(1962):1185–1204, 2012.
- [126] Timothy M Lenton, Johan Rockström, Owen Gaffney, Stefan Rahmstorf, Katherine Richardson, Will Steffen, and Hans Joachim Schellnhuber. Climate tipping points—too risky to bet against. Nature, 575(7784):592–595, 2019.
- [127] TM Lenton, VN Livina, Vasilis Dakos, and M Scheffer. Climate bifurcation during the last deglaciation? Climate of the Past, 8(4):1127–1139, 2012.
- [128] Simon A Levin. Ecosystems and the biosphere as complex adaptive systems. Ecosystems, 1(5):431–436, 1998.
- [129] Zhifang Li, Xiaojie Chen, Han-Xin Yang, and Attila Szolnoki. Game-theoretical approach for opinion dynamics on social networks. Chaos: An Interdisciplinary Journal of Nonlinear Science, 32(7), 2022.

- [130] Jianguo Liu, Thomas Dietz, Stephen R Carpenter, Marina Alberti, Carl Folke, Emilio Moran, Alice N Pell, Peter Deadman, Timothy Kratz, Jane Lubchenco, et al. Complexity of coupled human and natural systems. science, 317(5844):1513–1516, 2007.
- [131] Zhiguang Liu, Meng Gao, Zizhen Li, and Hua Liu. Dispersal, colored environmental noise, and spatial synchrony in population dynamics: analyzing a discrete host–parasitoid population model. Ecological research, 24:383–392, 2009.
- [132] Frida Lögdberg and Uno Wennergren. Spectral color, synchrony, and extinction risk. Theoretical Ecology, 5:545–554, 2012.
- [133] Edward N Lorenz. Deterministic nonperiodic flow 1. In Universality in Chaos, 2nd edition, pages 367–378. Routledge, 2017.
- [134] Alfred James Lotka. Elements of physical biology. Williams & Wilkins, 1925.
- [135] Wenjie Lu, Jiazheng Li, Yifan Li, Aijun Sun, and Jingyang Wang. A cnn-lstm-based model to forecast stock prices. Complexity, 2020(1):6622927, 2020.
- [136] Dingjun Luo. Bifurcation theory and methods of dynamical systems, volume 15. World Scientific, 1997.
- [137] Fatemehalsadat Madaeni, Karem Chokmani, Rachid Lhissou, Yves Gauthier, Simon Tolszczuk-Leclerc, et al. Convolutional neural network and long short-term memory models for ice-jam predictions. The Cryosphere, 16(4):1447–1468, 2022.
- [138] Muhammad Tariq Majeed and Aisha Tauqir. Effects of urbanization, industrialization, economic growth, energy consumption, financial development on carbon emissions: an extended stirpat model for heterogeneous income groups. Pakistan Journal of Commerce and Social Sciences (PJCSS), 14(3):652–681, 2020.
- [139] Benoit Mandelbrot et al. The variation of certain speculative prices. Journal of business, 36(4):394, 1963.
- [140] Sanna Markkanen and Annela Anger-Kraavi. Social impacts of climate change mitigation policies and their implications for inequality. Climate Policy, 19(7):827–844, 2019.
- [141] Luiz Marques and Luiz Marques. Climate feedbacks and tipping points. Capitalism and Environmental Collapse, pages 199–231, 2020.

- [142] Jerrold E Marsden and Marjorie McCracken. The Hopf bifurcation and its applications, volume 19. Springer Science & Business Media, 2012.
- [143] Mark D McDonnell and Lawrence M Ward. The benefits of noise in neural systems: bridging theory and experiment. Nature Reviews Neuroscience, 12(7):415–425, 2011.
- [144] David Armstrong McKay. How deforestation and climate change could push the amazon to a tipping point. Redirecting Flows-Navigating the Future of the Amazon, page 18.
- [145] Robert McSweeney. Explainer: Nine ‘tipping points’ that could be triggered by climate change. Carbon Brief, 10:191–203, 2020.
- [146] Donella H Meadows. Thinking in systems: A primer. Sustainability Institute, 2008.
- [147] J. Menard, T. M. Bury, C. T. Bauch, and M. Anand. When conflicts get heated, so does the planet: coupled social-climate dynamics under inequality. Proceedings of the Royal Society B: Biological Sciences, 288, 2021.
- [148] Jyler Menard, Thomas M Bury, Chris T Bauch, and Madhur Anand. When conflicts get heated, so does the planet: coupled social-climate dynamics under inequality. Proceedings of the Royal Society B, 288(1959):20211357, 2021.
- [149] Kimberley R Miner, Merritt R Turetsky, Edward Malina, Annett Bartsch, Johanna Tamminen, A David McGuire, Andreas Fix, Colm Sweeney, Clayton D Elder, and Charles E Miller. Permafrost carbon emissions in a changing arctic. Nature Reviews Earth & Environment, 3(1):55–67, 2022.
- [150] Frances C Moore, Katherine Lacasse, Katharine J Mach, Yoon Ah Shin, Louis J Gross, and Brian Beckage. Determinants of emissions pathways in the coupled climate–social system. Nature, 603(7899):103–111, 2022.
- [151] Matthias A Müller, Steffen Waldherr, and Frank Allgöwer. The transcritical bifurcation in absolutely stable feedback systems. In 2009 European Control Conference (ECC), pages 2146–2151. IEEE, 2009.
- [152] KE Muryshev, AV Eliseev, II Mokhov, and AV Timazhev. A lag between temperature and atmospheric co2 concentration based on a simple coupled model of climate and the carbon cycle. In Doklady earth sciences, volume 463, page 863. Citeseer, 2015.
- [153] Friedkin Noah and Johnsen Eugene. Social influence networks and opinion change. Advances in Group Processes, 16:1–29, 1999.

- [154] Karine Nyborg, John M Anderies, Astrid Dannenberg, Therese Lindahl, Caroline Schill, Maja Schlüter, W Neil Adger, Kenneth J Arrow, Scott Barrett, Stephen Carpenter, et al. Social norms as solutions. Science, 354(6308):42–43, 2016.
- [155] Paul E O’Keeffe and Sebastian Wieczorek. Tipping phenomena and points of no return in ecosystems: beyond classical bifurcations. SIAM Journal on Applied Dynamical Systems, 19(4):2371–2402, 2020.
- [156] Daniel Oro and Lúcia Freixas. Flickering body temperature anticipates criticality in hibernation dynamics. Royal Society Open Science, 8(1):201571, 2021.
- [157] Ilona M Otto, Jonathan F Donges, Roger Cremades, Avit Bhowmik, Richard J Hewitt, Wolfgang Lucht, Johan Rockström, Franziska Allerberger, Mark McCaffrey, Sylvanus SP Doe, et al. Social tipping dynamics for stabilizing earth’s climate by 2050. Proceedings of the National Academy of Sciences, 117(5):2354–2365, 2020.
- [158] Krishnendu Pal, Smita Deb, and Partha Sharathi Dutta. Tipping points in spatial ecosystems driven by short-range correlated noise. Physical Review E, 106(5):054412, 2022.
- [159] Shirin Panahi, Ling-Wei Kong, Mohammadamin Moradi, Zheng-Meng Zhai, Bryan Glaz, Mulugeta Haile, and Ying-Cheng Lai. Machine learning prediction of tipping in complex dynamical systems. Physical Review Research, 6(4):043194, 2024.
- [160] Minal Pathak, Shaurya Patel, and Shreya Some. Climate change mitigation and sustainable development goals: Evidence and research gaps. PLOS Climate, 3(3):e0000366, 2024.
- [161] Induja Pavithran and RI Sujith. Effect of rate of change of parameter on early warning signals for critical transitions. Chaos: An Interdisciplinary Journal of Nonlinear Science, 31(1), 2021.
- [162] Antonio F Peralta, János Kertész, and Gerardo Iñiguez. Opinion dynamics in social networks: From models to data. arXiv preprint arXiv:2201.01322, 2022.
- [163] Owen L Petchey, Andrew Gonzalez, and Howard B Wilson. Effects on population persistence: the interaction between environmental noise colour, intraspecific competition and space. Proceedings of the Royal Society of London. Series B: Biological Sciences, 264(1389):1841–1847, 1997.

- [164] Carly A Phillips, Astrid Caldas, Rachel Cleetus, Kristina A Dahl, Juan Declet-Barreto, Rachel Licker, L Delta Merner, J Pablo Ortiz-Partida, Alexandra L Phelan, Erika Spanger-Siegfried, et al. Compound climate risks in the covid-19 pandemic. Nature Climate Change, 10(7):586–588, 2020.
- [165] Łukasz Płociniczak. Hopf bifurcation in a conceptual climate model with ice–albedo and precipitation–temperature feedbacks. Nonlinear Analysis: Real World Applications, 51:102967, 2020.
- [166] Henri Poincaré. Les méthodes nouvelles de la mécanique céleste, volume 2. Gauthier-Villars et fils, imprimeurs-libraires, 1893.
- [167] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. Numerical recipes 3rd edition. Cambridge: New York, 2007.
- [168] Stefan Rahmstorf. Ocean circulation and climate during the past 120,000 years. Nature, 419(6903):207–214, 2002.
- [169] Hegselmann Rainer and Ulrich Krause. Opinion dynamics and bounded confidence: models, analysis and simulation. 2002.
- [170] David A Randall, Richard A Wood, Sandrine Bony, Robert Colman, Thierry Fichefet, John Fyfe, Vladimir Kattsov, Andrew Pitman, Jagadish Shukla, Jayaraman Srinivasan, et al. Climate models and their evaluation. In Climate change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR), pages 589–662. Cambridge University Press, 2007.
- [171] Nicholas G Reich, Justin Lessler, Krzysztof Sakrejda, Stephen A Lauer, Sopon Iamsirithaworn, and Derek AT Cummings. Case study in evaluating time series prediction models using the relative mean absolute error. The American Statistician, 70(3):285–292, 2016.
- [172] David N Reynolds and Francesco Tudisco. Unique nash equilibrium of a nonlinear model of opinion dynamics on networks with friction-inspired stubbornness. Nonlinearity, 38(5):055027, 2025.
- [173] Jörgen Ripa and Per Lundberg. Noise colour and the risk of population extinctions. Proceedings of the Royal Society of London. Series B: Biological Sciences, 263(1377):1751–1753, 1996.

- [174] Lasse Ruokolainen, Andreas Lindén, Veijo Kaitala, and Mike S Fowler. Ecological and evolutionary dynamics under coloured environmental variation. Trends in Ecology & Evolution, 24(10):555–563, 2009.
- [175] Chris Russill and Zoe Nyssa. The tipping point trend in climate change communication. Global environmental change, 19(3):336–344, 2009.
- [176] Mehrshad Sadria and Thomas M Bury. Fatenet: an integration of dynamical systems and deep learning for cell fate prediction. Bioinformatics, 40(9):btae525, 2024.
- [177] Tara N Sainath, Oriol Vinyals, Andrew Senior, and Haşim Sak. Convolutional, long short-term memory, fully connected deep neural networks. In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), pages 4580–4584. Ieee, 2015.
- [178] Athira Satheesh Kumar, Chris T Bauch, and Madhur Anand. Climate-denying rumor propagation in a coupled socio-climate model: Impact on average global temperature. PloS one, 20(1):e0317338, 2025.
- [179] Marten Scheffer. Ecology of shallow lakes, volume 1. Springer, 1998.
- [180] Marten Scheffer. Critical transitions in nature and society. Princeton University Press, 2009.
- [181] Marten Scheffer, Jordi Bascompte, William A Brock, Victor Brovkin, Stephen R Carpenter, Vasilis Dakos, Hermann Held, Egbert H Van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. Nature, 461(7260):53–59, 2009.
- [182] Marten Scheffer, Steve Carpenter, Jonathan A Foley, Carl Folke, and Brian Walker. Catastrophic shifts in ecosystems. Nature, 413(6856):591–596, 2001.
- [183] Marten Scheffer, Frances Westley, and William Brock. Slow response of societies to new problems: causes and costs. Ecosystems, 6(5):493–502, 2003.
- [184] Maja Schlueter, Ryan RJ Mcallister, Robert Arlinghaus, Nils Bunnefeld, Klaus Eisenack, Frank Hoelker, Eleanor J MILNER-GULLAND, Birgit Müller, Emily Nicholson, Martin Quaas, et al. New horizons for managing the environment: A review of coupled social-ecological systems modeling. Natural Resource Modeling, 25(1):219–272, 2012.

- [185] Nafiz Shahriar. “what is a convolutional neural network (cnn) – deep learning, 2025. Accessed: 2025-11-07.
- [186] Yoon Ah Shin, Sara M Constantino, Brian Beckage, and Katherine Lacasse. Climate change and opinion dynamics models: Linking individual, social, and institutional level changes. Current Opinion in Behavioral Sciences, 64:101528, 2025.
- [187] Longmei Shu and Feng Fu. Determinants of successful mitigation in coupled social-climate dynamics. Proceedings of the Royal Society A, 479(2280):20230679, 2023.
- [188] Ram Sigdel, Madhur Anand, and Chris T Bauch. Convergence of socio-ecological dynamics in disparate ecological systems under strong coupling to human social systems. Theoretical Ecology, 12(3):285–296, 2019.
- [189] Ram P Sigdel, Madhur Anand, and Chris T Bauch. Competition between injunctive social norms and conservation priorities gives rise to complex dynamics in a model of forest growth and opinion dynamics. Journal of theoretical biology, 432:132–140, 2017.
- [190] Mannava VK Sivakumar and Robert Stefanski. Climate and land degradation—an overview. Climate and land degradation, pages 105–135, 2007.
- [191] Didier Sornette. Why stock markets crash: critical events in complex financial systems. Princeton university press, 2017.
- [192] Tommaso Spanio, Jorge Hidalgo, and Miguel A Muñoz. Impact of environmental colored noise in single-species population dynamics. Physical Review E, 96(4):042301, 2017.
- [193] Steven H Strogatz. Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. Chapman and Hall/CRC, 2024.
- [194] Alessandro Tavoni, Maja Schlüter, and Simon Levin. The survival of the conformist: social pressure and renewable resource management. Journal of theoretical biology, 299:152–161, 2012.
- [195] KC Taylor, GW Lamorey, GA Doyle, RB Alley, PM Grootes, Paul A Mayewski, JWC White, and LK Barlow. The ‘flickering switch’ of late pleistocene climate change. Nature, 361(6411):432–436, 1993.

- [196] Vivek A Thampi, Madhur Anand, and Chris T Bauch. Socio-ecological dynamics of caribbean coral reef ecosystems and conservation opinion propagation. Scientific reports, 8(1):2597, 2018.
- [197] Jonathan D Touboul, Charlotte Piette, Laurent Venance, and G Bard Ermentrout. Noise-induced synchronization and antiresonance in interacting excitable systems: applications to deep brain stimulation in parkinson’s disease. Physical Review X, 10(1):011073, 2020.
- [198] Cyril Touzé. Normal form theory and nonlinear normal modes: theoretical settings and applications. In Modal analysis of nonlinear mechanical systems, pages 75–160. Springer, 2014.
- [199] Martin H Trauth, Asfawossen Asrat, Markus L Fischer, Peter O Hopcroft, Verena Foerster, Stefanie Kaboth-Bahr, Karin Kindermann, Henry F Lamb, Norbert Marwan, Mark A Maslin, et al. Early warning signals of the termination of the african humid period (s). Nature Communications, 15(1):3697, 2024.
- [200] Jorge R Tredicce, Gian Luca Lippi, Paul Mandel, Basile Charasse, Aude Chevalier, and B Picqué. Critical slowing down at a bifurcation. American Journal of Physics, 72(6):799–809, 2004.
- [201] Merritt R Turetsky, Benjamin W Abbott, Miriam C Jones, Katey Walter Anthony, David Olefeldt, Edward AG Schuur, Guido Grosse, Peter Kuhry, Gustaf Hugelius, Charles Koven, et al. Carbon release through abrupt permafrost thaw. Nature Geoscience, 13(2):138–143, 2020.
- [202] Martijn van de Pol, Yngvild Vindenes, Bernt-Erik Sæther, Steinar Engen, Bruno J Ens, Kees Oosterbeek, and Joost M Tinbergen. Poor environmental tracking can make extinction risk insensitive to the colour of environmental noise. Proceedings of the Royal Society B: Biological Sciences, 278(1725):3713–3722, 2011.
- [203] Nicolaas Godfried Van Kampen. Stochastic processes in physics and chemistry, volume 1. Elsevier, 1992.
- [204] David A Vasseur. Populations embedded in trophic communities respond differently to coloured environmental noise. Theoretical Population Biology, 72(2):186–196, 2007.

- [205] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel type of phase transition in a system of self-driven particles. Physical review letters, 75(6):1226, 1995.
- [206] K Walter Anthony, T Schneider von Deimling, I Nitze, S Frolking, A Emond, R Daanen, P Anthony, P Lindgren, B Jones, and G Grosse. 21st-century modeled permafrost carbon emissions accelerated by abrupt thaw beneath lakes, *nat. commun.*, 9, 3262, 2018.
- [207] Rong Wang, John A Dearing, Peter G Langdon, Enlou Zhang, Xiangdong Yang, Vasilis Dakos, and Marten Scheffer. Flickering gives early warning signals of a critical transition to a eutrophic lake state. Nature, 492(7429):419–422, 2012.
- [208] Zhen Wang, Chris T Bauch, Samit Bhattacharyya, Alberto d’Onofrio, Piero Manfredi, Matjaž Perc, Nicola Perra, Marcel Salathé, and Dawei Zhao. Statistical physics of vaccination. Physics Reports, 664:1–113, 2016.
- [209] Zhen Wang, Chris T. Bauch, Samit Bhattacharyya, Alberto d’Onofrio, Piero Manfredi, Matjaž Perc, Nicola Perra, Marcel Salathé, and Dawei Zhao. Statistical physics of vaccination. Physics Reports, 664:1–113, 2016. Statistical physics of vaccination.
- [210] Zhen Wang, Marko Jusup, Rui-Wu Wang, Lei Shi, Yoh Iwasa, Yamir Moreno, and Jürgen Kurths. Onymity promotes cooperation in social dilemma experiments. Science advances, 3(3):e1601444, 2017.
- [211] Zhen Wang, Chunjiang Mu, Shuyue Hu, Chen Chu, and Xuelong Li. Modelling the dynamics of regret minimization in large agent populations: a master equation approach. In IJCAI, pages 534–540, 2022.
- [212] Lawrence M Ward, Shannon E MacLean, and Aaron Kirschner. Stochastic resonance modulates neural synchronization within and between cortical sources. PloS one, 5(12):e14371, 2010.
- [213] Richard G Williams, Anna Katavouta, and Philip Goodwin. Carbon-cycle feedbacks operating in the climate system. Current Climate Change Reports, 5(4):282–295, 2019.
- [214] MS Williamsona, TM Lentona, JG Shepherd, and NR Edwardsd. An efficient numerical terrestrial scheme (ents) for earth system modeling. Ecol Model, 198:362–374, 2006.

- [215] C Wissel. A universal law of the characteristic return time near thresholds. Oecologia, 65(1):101–107, 1984.
- [216] Peter John Wood. Climate change and game theory. Annals of the New York Academy of Sciences, 1219(1):153–170, 2011.
- [217] Nico Wunderling, Ricarda Winkelmann, Johan Rockström, Sina Loriani, David I Armstrong McKay, Paul DL Ritchie, Boris Sakschewski, and Jonathan F Donges. Global warming overshoots increase risks of climate tipping cascades in a network model. Nature Climate Change, 13(1):75–82, 2023.
- [218] Tianjun Zhou. New physical science behind climate change: What does ipcc ar6 tell us? The Innovation, 2(4), 2021.
- [219] Xiaotian Zhou, Haoxin Sun, Wanyue Xu, Wei Li, and Zhongzhi Zhang. Friedkin-johnsen model for opinion dynamics on signed graphs. IEEE Transactions on Knowledge and Data Engineering, 2024.
- [220] Jiefan Zhu, Yiping Yao, Wenjie Tang, and Haoming Zhang. An agent-based model of opinion dynamics with attitude-hiding behaviors. Physica A: Statistical Mechanics and its Applications, 603:127662, 2022.

APPENDICES

Appendix A

Supporting Information for chapter 2

A.1 Deep Learning Model Architecture

The deep learning architecture used in this study was implemented as a Sequential model using the Keras API, designed for multi-class sequence classification. Initial experiments revealed that a single one-dimensional convolutional (Conv1D) layer was insufficient for the model to learn the necessary temporal features and bifurcation-specific patterns. To address this limitation, we added a second convolutional layer, which significantly improved feature extraction and overall classification performance. The final architecture begins with two Conv1D layers: the first comprising 50 filters and the second 100 filters, both using a kernel size of 10, ReLU activation, and ‘same’ padding to preserve the temporal dimension of the input. These convolutional layers are followed by a Dropout layer with a rate of 0.05 to mitigate overfitting. A MaxPooling1D layer with a pool size of 2 and stride of 2 is then applied to downsample the feature maps.

The convolutional block is followed by two standard, stacked unidirectional Long Short-Term Memory (LSTM) layers. The first LSTM layer includes 50 memory cells and returns sequences to preserve temporal structure for the next LSTM layer, which contains 10 units. Each LSTM layer is followed by an additional Dropout layer with a rate of 0.05. The network concludes with a fully connected Dense output layer with 4 units, corresponding to the number of classes, and uses a softmax activation function to output class probabilities. All layers utilize the LeCun kernel initializer for weight initialization.

For training, the model uses the Adam optimizer with a learning rate of 0.01. Training is conducted for up to 200 epochs with a batch size of 128. An Early Stopping callback is employed to halt training if the validation accuracy does not improve for 10 consecutive epochs, and a Model Checkpoint is configured to save the best-performing model based on

validation accuracy. The loss function used is 'sparse categorical crossentropy', suitable for multi-class classification tasks with integer labels. Model performance is monitored using both 'accuracy' and 'sparse categorical accuracy' metrics during training.

After experimenting with CNN layers, LSTM layers, and Inception layers, we found that the combination of CNN and LSTM layers produced the best results. We also implemented a simplified LSTM-only model to perform an ablation study. The LSTM-only architecture consisted of two stacked LSTM layers (50 and 10 units) with the same dropout, dense output, and training procedure as the CNN–LSTM model. All models were trained using the same dataset and evaluated under consistent conditions. Our choice of architecture and hyperparameters was informed by prior work by [34], and we further refined these settings using a grid search algorithm around the values suggested in that study to ensure optimal performance. We used the pre-processed time series for each redness value, allocating 90% of the time series for the training process and 10% for validation. Figure A.1 provides a detailed overview of the deep learning (DL) architecture used in this study. The learning rate was set to 0.0005, and training was done in 200 epochs, and early stopping was employed to prevent overfitting. The code is written in python using TensorFlow 2.12.1.

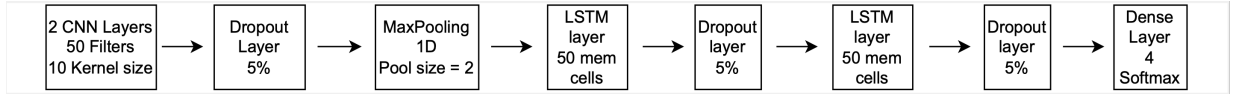


Figure A.1: Deep learning architecture

A.2 Validation on Bifurcating and Null Dynamical Systems

To evaluate the model's ability to avoid false positives and correctly assign low confidence to bifurcation classes when no critical transition is present, we simulate time series from six nonlinear dynamical systems. In each case, control parameters are held fixed such that the system remains outside any bifurcation regime throughout the simulation. These represent purely null dynamics, meaning no tipping point occurs. This setup allows us to assess whether the classifier can reliably withhold predictions of bifurcations when presented with stable, non-transitioning systems, an essential capability for reducing false alarms in real-world applications.

Bifurcating Systems:

- **Rosenzweig–MacArthur Predator–Prey Model (Hopf or Transcritical):**

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{k}\right) - \frac{axy}{1 + ahx}, \quad \frac{dy}{dt} = \frac{eaxy}{1 + ahx} - my$$

with fixed parameters $r = 4$, $k = 1.7$, $e = 0.5$, $h = 0.15$, $m = 2$. For the transcritical bifurcation, the predation rate a is held constant at 6, and for hopf is held constant at 12.

- **Fold Bifurcation Model (Harvesting):**

$$\frac{dx}{dt} = x(1 - x) - h \cdot \frac{x^2}{x^2 + 0.01}$$

where the harvesting rate h is held constant at 0.1 over the course of the simulation.

- **Van der Pol Oscillator (Hopf):**

$$\frac{dx}{dt} = y, \quad \frac{dy}{dt} = \mu(1 - x^2)y - x$$

with $\mu = -7.0$.

- **SIS Epidemic Model (Transcritical):**

$$\frac{dI}{dt} = \beta I(1 - I) - \gamma I$$

where $\beta = 4$, $\gamma = 5$.

- **Neural Activation Model (Fold):**

$$\frac{du}{dt} = -u + \frac{r}{1 + \exp(-5(u - 1))}$$

with $r = 6$.

Null Dynamics: The null simulations are derived by fixing the control parameters in each system, preventing any bifurcation. For example, in the Rosenzweig–MacArthur model, $a = 6$ is held constant to avoid crossing a bifurcation threshold. Similarly, in the fold and transcritical systems, the control parameters are set below the bifurcation point and remain unchanged throughout the simulation. The results are presented in Figures [A.2](#) and [A.3](#).

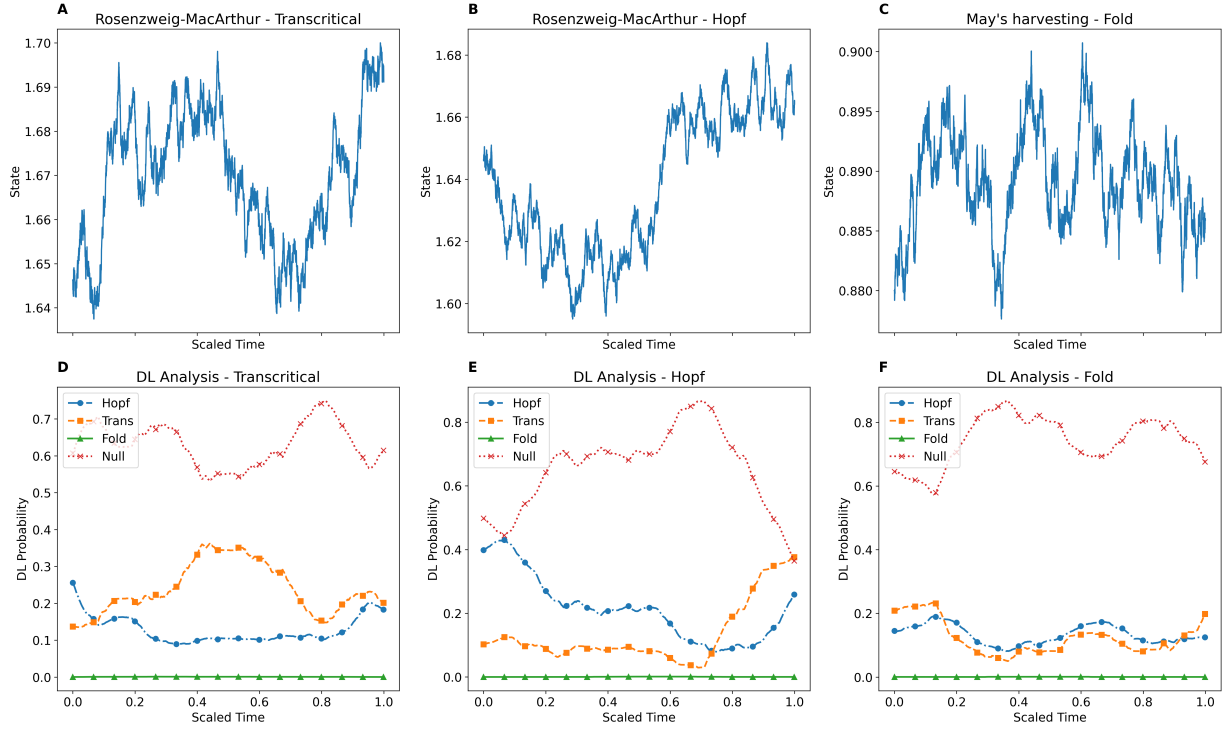


Figure A.2: **(A–C)** Synthetic null time series from three canonical bifurcation models with additive white noise all far from the bifurcation point: (A) Rosenzweig–MacArthur predator-prey system (B) Rosenzweig–MacArthur, (C) May’s harvesting model **(D–F)** Deep learning model predictions for the systems in (A–C), showing temporal evolution of bifurcation class probabilities. Probabilities are represented with distinguishable markers and line styles for print clarity: triangle (fold), square (transcritical), circle (Hopf), and x (null). In all subplots, the null class consistently exhibits the highest predicted probability compared to the other bifurcation types.

As shown, the DL model predominantly assigns high probabilities to the Null class across all null simulations, correctly identifying the absence of bifurcation dynamics. Minor decreases in Null probability observed in some cases can be interpreted either as small classification errors, reflecting the imperfect nature of any learned model, or as the model responding to residual structure in the time series that deviates slightly from purely stochastic noise.

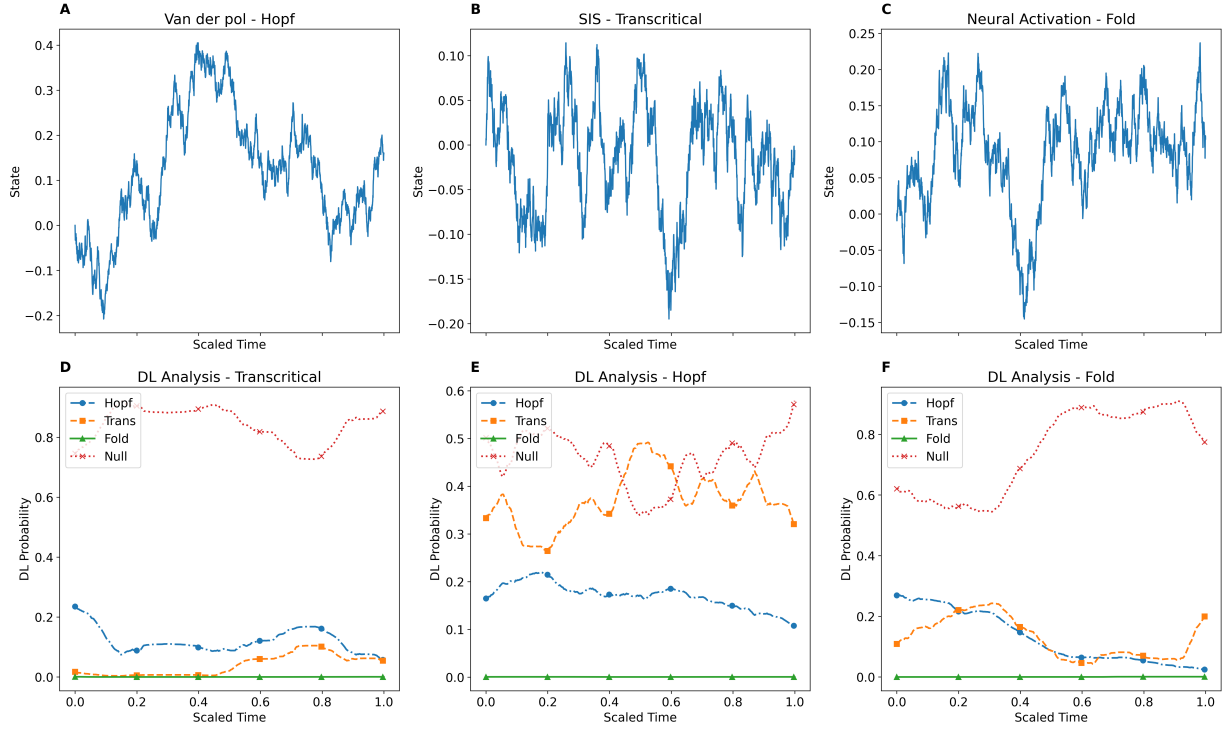


Figure A.3: **(A–C)** Synthetic null time series from three canonical bifurcation models with additive white noise all far from the bifurcation point: (A) Van der pol system, (B) SIS model, (C) Neural Activation model **(D–F)** Deep learning model predictions for the systems in (A–C), showing temporal evolution of bifurcation class probabilities. Probabilities are represented with distinguishable markers and line styles for print clarity: triangle (fold), square (transcritical), circle (Hopf), and x (null). The null class consistently exhibits the highest predicted probability in all subplots compared to the other bifurcation types.

Appendix B

Supporting Information for chapter 3

B.1 Deep Learning Architecture

The deep learning model used in this study comprises three main components: convolutional feature extraction, temporal sequence modelling, and classification. The model was implemented in Keras using the TensorFlow backend, and designed to capture both spatial patterns in the input time series and their temporal evolution.

Feature Extraction: To extract localized patterns across long temporal windows, the model starts with two 1D convolutional layers using large kernels. Specifically:

- **Conv1D layer 1:** 50 filters, kernel size = 300, activation = ReLU, padding = 'same'
- **Conv1D layer 2:** 100 filters, kernel size = 300, activation = ReLU, padding = 'same'
- **Dropout:** dropout rate = 0.05
- **MaxPooling1D:** pool size = 2, stride = 2, padding = 'valid'

Temporal Modelling: The temporally encoded features are passed through two stacked LSTM layers:

- **LSTM layer 1:** 50 memory cells, return_sequences = True
- **Dropout:** dropout rate = 0.05
- **LSTM layer 2:** 10 memory cells

- **Dropout:** dropout rate = 0.05

Classification: The final hidden representation is passed to a fully connected output layer:

- **Dense layer:** 2 units, activation = softmax

This outputs the class probabilities for binary classification.

Training Configuration: The model was compiled with the Adam optimizer with a learning rate of 0.01 and trained using the sparse categorical cross-entropy loss. A batch size of 32 was used. To prevent overfitting, early stopping with a patience of 2 epochs and model checkpointing based on validation accuracy were applied. The model was trained for up to 5 epochs, and the best-performing model (based on validation accuracy) was saved. The input shape for each sample was (5000, 2), corresponding to 5000 time steps with 2 features per time step.

This configuration enabled the model to achieve more than 95% training accuracy within 5 epochs on the given training data.

B.2 Data Availability

All the codes and empirical data used in this research can be found in the following repository:

<https://github.com/Yazdan-Babazadeh/Echoes-Before-Collapse>

Appendix C

Supporting Information for chapter 4

C.1 Stability Analysis of the social model

The social model is given by:

$$\frac{dx}{dt} = kx(1-x)(-\beta + f(T) + \delta(2x-1)) \quad (\text{C.1})$$

To simplify the analysis, we define $-\beta + f(T)$ as ψ . This model has two equilibrium points at 0 and 1 and also another equilibrium point exists if:

$$\psi + \delta(2x-1) = 0 \quad (\text{C.2})$$

This equilibrium point happens in the range of (0,1) if

$$-\delta < \psi < \delta \quad (\text{C.3})$$

To evaluate the stability of the equilibrium points, we take the second derivative of x with respect to time:

$$k\psi + k\delta(4x-1) - 2k\psi x - k\delta(6x^2-2x) \quad (\text{C.4})$$

For $x = 0$, this becomes:

$$k(\psi - \delta) \quad (\text{C.5})$$

So if $\psi > \delta$ it destabilises $x=0$.

For $x = 1$, this becomes:

$$-k(\psi + \delta) \tag{C.6}$$

if $\psi < -\delta$, $x = 1$ will be unstable.

As a result of a positive and high value of δ , if $\psi < \delta$, then three equilibrium points exist at $x = 0$, $x = 1$, and one more in $[0, 1]$. In this case, $x = 0$ and $x = 1$ are stable, and the third equilibrium point is unstable. If the climate system goes through a tipping point and a sudden increase in temperature, $f(T)$ will increase, leading to an increase in ψ . Due to this increase, if $\psi > \delta$, the third equilibrium point collides with $x = 0$ and gets annihilated. Additionally, $x = 0$ becomes unstable while $x = 1$ remains a stable equilibrium point. As a result, the social model converges to $x = 1$ for $\psi > \delta$ and to $x = 0$ for $\psi < \delta$, as the initial value of x at the beginning of the simulation is very close to zero.

C.2 Sensitivity of Climate Impact to Social Learning Rate

To better understand the observed non-monotonic relationship between social learning rate κ and climate tipping dynamics, we performed a local sensitivity analysis focused on the cumulative climate impact, measured by the area under the temperature anomaly curve (AUC). Specifically, we computed the derivative of the AUC difference (between the modified and baseline models) with respect to κ across a range of tipping intensities $R_{\max}$. Both the signed and absolute sensitivities were examined to capture the direction and magnitude of the response.

Figure C.1 presents the results of this analysis. The left panel shows $\partial\text{AUC}/\partial\kappa$ as a function of κ , revealing strong positive sensitivity at low values of κ , which then decays and oscillates around zero. The right panel displays the absolute sensitivity $|\partial\text{AUC}/\partial\kappa|$, highlighting a peak in sensitivity at intermediate κ values for all examined $R_{\max}$ levels.

These results suggest that the system is most sensitive to changes in social learning when κ falls in extreme ranges. At very low κ , social behavior adjusts too slowly to influence emissions before the tipping point is triggered. At very high κ , the system rapidly locks into mitigation behavior, thereby reducing variability. In the intermediate regime, the alignment between social transition timing and the climate tipping window amplifies system response, making outcomes highly sensitive to small changes in κ .

This analysis reveals a critical vulnerability in the social-climate system: moderate social responsiveness may maximize tipping risk due to the delayed coordination of collective action. These findings underscore the importance of understanding not only the presence of tipping dynamics but also their interaction with the timescales of social processes.

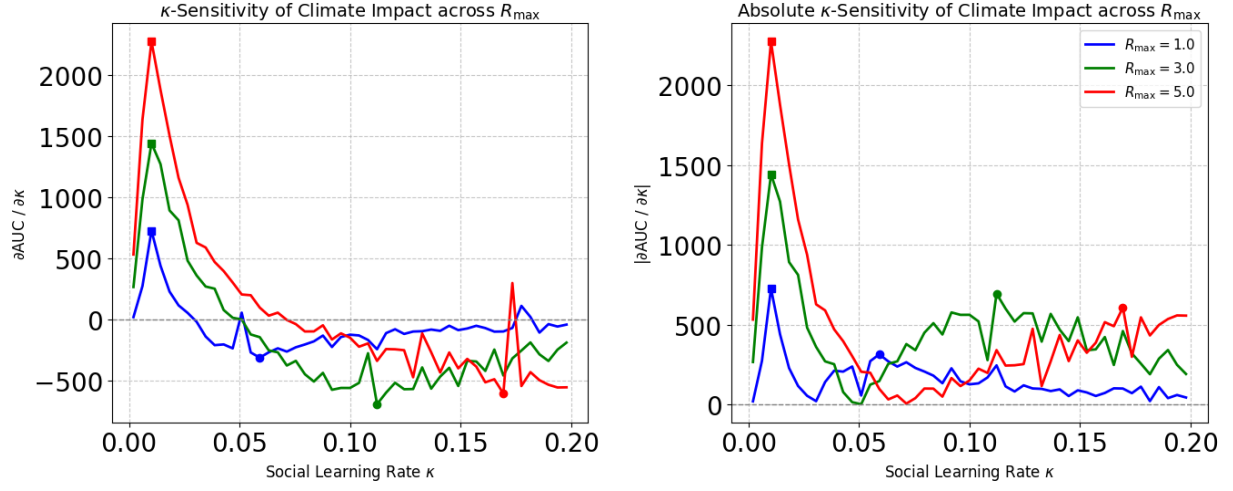


Figure C.1: Sensitivity of climate impact to the social learning rate κ across different tipping strengths $R_{\max}$. Left: Signed derivative $\partial \text{AUC} / \partial \kappa$. Right: Absolute value $|\partial \text{AUC} / \partial \kappa|$. Peak sensitivity occurs in an intermediate regime of κ , particularly when $R_{\max}$ is high.

C.3 Table. Labels for state variables and dynamic processes in the socio-climate model.

Parameter	Definition	Baseline Values / Intervals	Unit
κ	Social Learning Rate	(0, 0.2)	yr ⁻¹
δ	Strength of Social Norm	(0,2)	1
β	Net Cost of Mitigation	(0,2)	1
f_{max}	Maximum of warming cost function $f(T)$	(5)	1
w	Nonlinearity of $f(T)$	(3)	K ⁻¹
T_{crit}	Critical temperature of $f(T)$	(1.5)	K
R_{max}	Maximum of function $R(T)$	(0,5)	GtC/year
R_0	Nonlinearity of $R(T)$	(5)	K ⁻¹
T_c	Critical temperature $R(T)$	(2,3)	K
x_0	Initial proportion of mitigators	(0.05)	1
T_R	Temperature of freezing point for water	273.15	K
σ	Stefan–Boltzmann constant	5.67×10^{-8}	W m ⁻² K ⁻⁴
C_{at0}	Initial CO ₂ in atmosphere	596	GtC
C_{ao0}	Initial CO ₂ in ocean reservoir	1.5×10^5	GtC
C_{veg0}	Initial CO ₂ in vegetation reservoir	550	GtC
C_{so0}	Initial CO ₂ in soil reservoir	1500	GtC
k_p	Photosynthesis rate constant	0.184	yr ⁻¹
k_{MM}	Photosynthesis normalising constant	1.478	1
k_c	Photosynthesis compensation point	29×10^{-6}	1
k_M	Half-saturation point for photosynthesis	120×10^{-6}	1
k_a	Mole volume of atmosphere	1.773×10^{20}	moles
k_r	Plant respiration constant	0.092	yr ⁻¹
k_A	Plant respiration normalising constant	8.7039×10^9	1
E_a	Plant respiration activation energy	54.83	J mol ⁻¹
k_{sr}	Soil respiration rate constant	0.034	yr ⁻¹
k_B	Soil respiration normalising constant	157.072	1
k_t	Turnover rate constant	0.092	yr ⁻¹
c	Specific heat capacity of Earth's surface	4.69×10^{23}	J K ⁻¹
a_E	Earth's surface area	5.101×10^{14}	m ²

Parameter	Definition	Baseline Values / Intervals	Unit
L	Latent heat per mole of water	43655	J mol^{-1}
R	Molar gas constant	8.314	$\text{J mol}^{-1} \text{K}^{-1}$
H	Relative humidity	0.5915	1
A	Surface albedo	0.225	1
S	Solar flux	1368	W m^{-2}
$\tau(\text{CH}_4)$	Methane opacity	0.0231	1
P_0	Water vapor saturation constant	1.4×10^{11}	Pa
F_0	Ocean flux rate constant	2.5×10^{-2}	yr^{-1}
χ	Characteristic CO_2 solubility	0.3	1
ζ	Evasion factor	50	1
s	Half-saturation time for $\epsilon(t)$ from 2014	50	yr
ϵ_{max}	Maximum change in $\epsilon(t)$ from 2014	7	GtC yr^{-1}

C.4 Sensitivity Analysis

We have performed a sensitivity analysis to understand the influence of each parameter on the results provided in the previous figures. Figure C.2 shows the sensitivity analysis result in a tornado plot for the difference in area under the curve and the peak of temperature anomaly. We have varied all the parameters by 5 percent of their baseline values to generate these plots. By observing both of the plots we can see that Solar flux, initial temperature, Albedo, and Humidity level have the highest influence on the value of the results we have reported. This analysis, also reveals the non-monotonic relation between the Difference in AUC and the peak of temperature against the parameters. For instance, an increase in solar flux will increase the difference in AUC while the same result can be achieved by decreasing the Albedo. Additionally, we observe that the sensitivity values are not symmetric, meaning that the constant variation of parameters results in different amounts of change in the difference in AUC and the peak of the temperature anomaly.

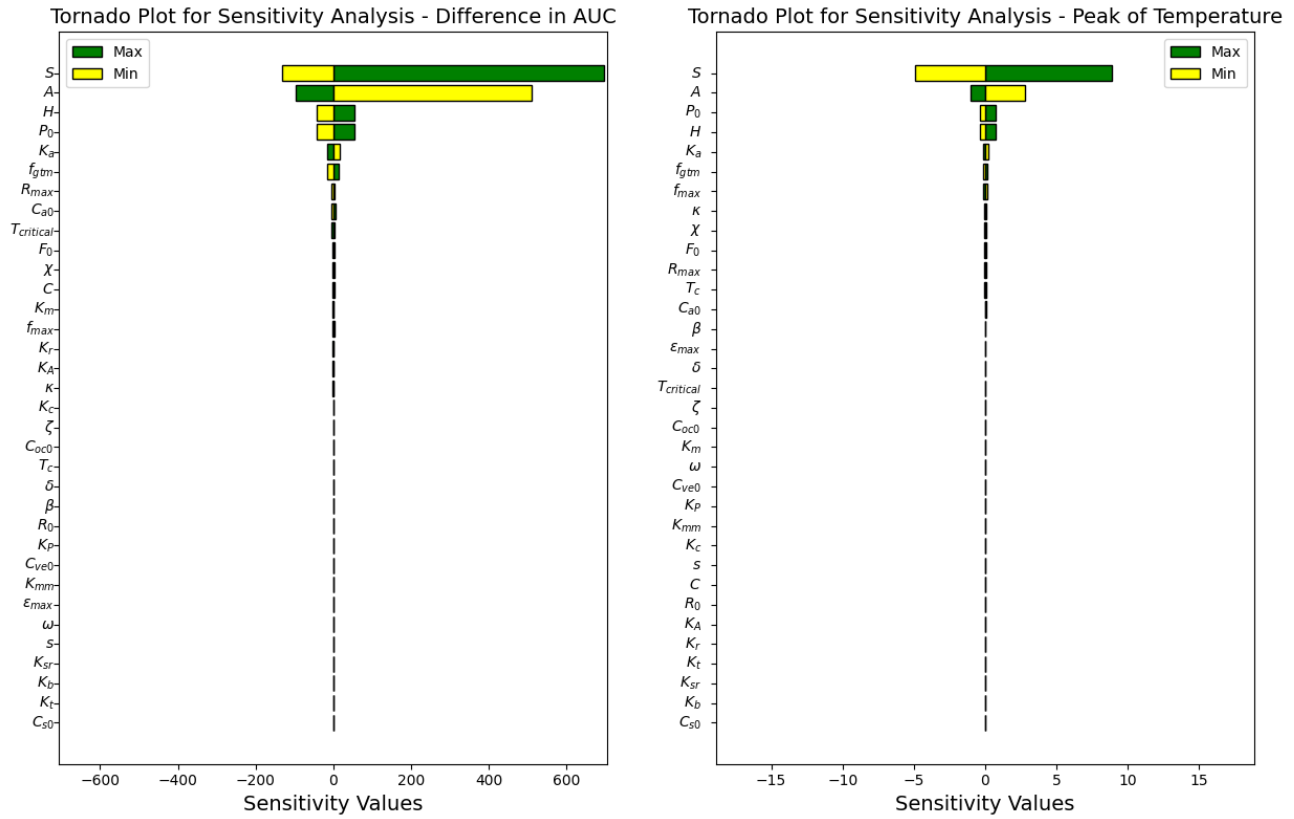


Figure C.2: Tornado plot showing the sensitivity analysis for the Difference in the area under the curve (left) and the peak of temperature (right). The yellow (green) bar shows the sensitivity values when the specified parameter has its upper (lower) bound value. The definitions of all parameters and variables are provided in a table in 5.2.

C.5 Historical record of CO_2 emission

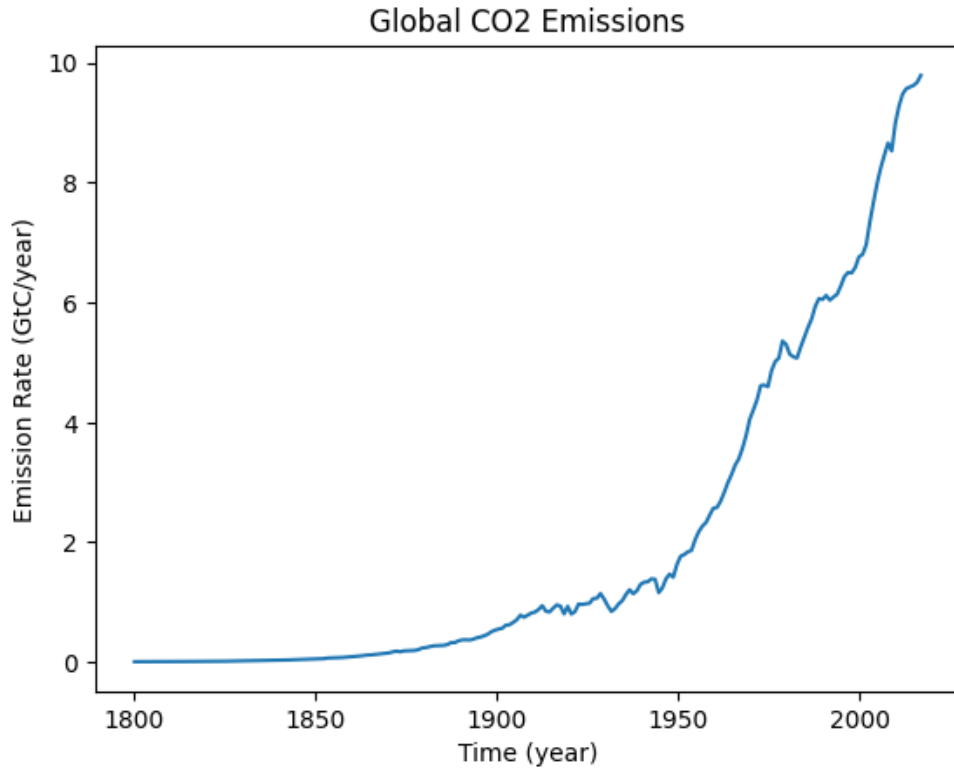


Figure C.3: Historical record for Emission rate Vs Time is plotted. This data has been gathered from 1800 to 2017.

C.6 Data and Code Availability

All data and code supporting this study are available at: [GitHub Repository](https://github.com/Yazdan-Babazadeh/Socio-Climate-model.git).
<https://github.com/Yazdan-Babazadeh/Socio-Climate-model.git>

Appendix D

Supporting Information for chapter 5

D.1 Formulation of Climate–Carbon Model

We have used the Earth System Model [122] combined with reduced ocean dynamics [152]. The full earth system model is as follows:

$$\frac{dC_{at}}{dt} = \epsilon(t) - P + R_{veg} + R_{so} - F_{oc} \quad (D.1)$$

$$\frac{dC_{oc}}{dt} = F_{oc} \quad (D.2)$$

$$\frac{dC_{veg}}{dt} = P - R_{veg} - L \quad (D.3)$$

$$\frac{dC_{so}}{dt} = L - R_{so} \quad (D.4)$$

$$c \frac{dT}{dt} = (F_d - \sigma T^4) a_E \quad (D.5)$$

P represents photosynthesis which takes the following form:

$$P(C_{at}, T) = k_p C_{v0} k_{MM} \left(\frac{pCO_{2a} - k_c}{K_M + pCO_{2a} - k_c} \right) \left(\frac{(15 + T)^2 (25 - T)}{5625} \right) \quad (D.6)$$

for $pCO_{2a} \geq k_c$ and $-15 \leq T \leq 25$, and zero otherwise. pCO_{2a} is defined as the ratio of moles of CO_2 in the atmosphere to the total number of moles of molecules in the atmosphere k_a :

$$pCO_{2a} = \frac{f_{gtm}(C_{at} + C_{at0})}{k_a} \quad (D.7)$$

where $f_{gtm} = 8.3259 * 10^{13}$ is the conversion factor from GtC to moles of carbon and C_{at0} is the initial level of CO_2 in the atmosphere.

Plant respiration takes the form:

$$R_{veg}(T, C_{veg}) = k_r C_{veg} k_A e^{\frac{-E_a}{R(T+T_0)}} \quad (D.8)$$

Soil respiration takes the form :

$$R_{so}(T, C_{so}) = k_{sr} C_{so} k_B e^{\frac{-308.56}{T+T_0-227.13}} \quad (D.9)$$

Turnover (constant fraction of plants dying in a given unit of time) takes the form:

$$L(C_{veg}) = k_t C_{veg} \quad (D.10)$$

flux of CO_2 from the atmosphere to the ocean takes the form

$$F_{oc}(C_{at}, C_{oc}) = F_{0\chi}(C_{at} - \zeta \frac{C_{at0}}{C_{oc0} C_{oc}}) \quad (D.11)$$

where χ is the characteristic solubility of CO_2 in water and ζ is the evasion factor.

The net downward flux of absorbed radiation at the surface is :

$$F_d = \frac{(1-A)S}{4} (1 + \frac{3\tau}{4}) \quad (D.12)$$

where A is the surface albedo, S is the incoming solar flux and τ is the vertical opacity of the greenhouse atmosphere. The opacity of each greenhouse is given by:

$$\tau(CO_2) = 1.73(pCO_2)^{0.263} \quad (D.13)$$

$$\tau(H_2O) = 0.0126(HP_0 e^{\frac{-L}{RT}})^{0.503} \quad (D.14)$$

$$\tau(CH_4) = 0.0231 \quad (D.15)$$

H is the relative humidity, P_0 is the water vapor saturation constant, L is the latent heat per mole of water, and R is the molar gas constant.

D.2 Full details of the replicator–climate–carbon model coupling

The coupled system is solved numerically using the stiff solver BDF within `solve_ivp`. The initial condition sets carbon pools to zero perturbation, and social shares are initialized

such that on average only 5% of the population is mitigating at the start of the simulation. Parameter sweeps are performed over four key social parameters:

$$\kappa, \quad \alpha_{\max}, \quad \delta, \quad m_{\max},$$

with both low and high critical temperature thresholds ($T_c = 2$ and $T_c = 3$).

Table D.1 lists all constants and baseline values used in the simulation, divided into carbon–climate subsystem and social subsystem. Where parameters are varied in sweeps, the range is reported alongside the baseline.

D.3 Full details of the FJ social model–forest cover coupling

We first allow the forest subsystem to evolve in isolation. From $F(0) = 0.2$, we switch off social interactions and evolve only the forest equation up to $t_{\text{on}} = 400$. This pre-activation phase allows the ecological state to relax toward its intrinsic trajectory without social pressure. At t_{on} we switch on the FJ social layer and let opinions push back on the forest. To keep opinions physical, we confine $X_i(t)$ to $[-1, 1]$ by clipping states at every right-hand-side evaluation and zeroing any outward-pointing velocities at the bounds.

We explore two social initializations that mirror our binary vs. continuous comparison: a continuous mode (all $X_i(0) = 0$) and a binary-proxy mode (half $X_i(0) = -1$, half $+1$, shuffled). After activation, we let the coupled system run until the forest response stabilizes; concretely, we declare convergence once $|\dot{F}(t)| < \text{tolF} = 10^{-6}$ holds after a short buffer ($t \geq t_{\text{on}} + 50$) so that transient switching effects are excluded. Integration uses the stiff BDF solver with a uniform evaluation grid of ≈ 50 points per time unit, up to $\text{max time} = 5000$.

For each parameter setting we simulate both modes, and quantify how much the two behaviors differ only in the post-activation window. We do so via phase-2 relative integrated absolute error (RIAE) metrics for forest cover F and the population mean $\langle X \rangle$. We repeat this procedure across two runaway-threshold scenarios $F_c \in \{0.2, 0.8\}$ and perform one-dimensional sweeps over $R_{\max}$, Ψ , A , λ_0 , c , and v , keeping all other parameters at their baselines (Table D.3). Each point aggregates 20 stochastic replicates (independent Λ_i draws with fixed seeds), and we report means with standard errors.

Table D.1: Parameters of the replicator–climate–carbon coupled model (carbon and climate subsystem).

Parameter	Value / Range	Description
<i>Initial carbon pools</i>		
C_{s0}	1500	Soil carbon initial (GtC)
C_{ve0}	550	Vegetation carbon initial (GtC)
C_{a0}	596	Atmospheric carbon initial (GtC)
C_{oc0}	1.5×10^5	Ocean carbon initial (GtC)
<i>Carbon–climate parameters</i>		
f_{gtm}	8.3259×10^{13}	Gas transfer constant
K_a	1.773×10^{20}	Carbonate equilibrium constant
K_p	0.184	Photosynthesis prefactor
K_A	8.7039×10^9	Vegetation growth factor
K_{mm}	1.478	Michaelis–Menten coefficient
K_c	29×10^{-6}	Compensation concentration
K_m	120×10^{-6}	Michaelis constant
T_0	288.15	Reference absolute temperature (K)
K_r	0.092	Vegetation respiration constant
E_a	54.83	Activation energy (kJ mol ^{−1})
R	8.314	Universal gas constant (J mol ^{−1} K ^{−1})
K_{sr}	0.034	Soil respiration constant
K_b	157.072	Soil carbon feedback factor
K_t	0.092	Vegetation litterfall rate
F_0	2.5×10^{-2}	Ocean flux baseline
ξ	0.3	Ocean exchange coefficient
ζ	50	Ocean buffer factor
H	0.5915	Humidity factor
P_0	1.4×10^{11}	Reference pressure constant
Latent	43655	Latent heat constant
A	0.225	Planetary albedo
S	1368	Solar constant (W m ^{−2})
capacity	4.69×10^{23}	Heat capacity of climate system (J K ^{−1})
σ	5.67×10^{-8}	Stefan–Boltzmann constant
a_E	5.101×10^{14}	Effective emitting area (m ²)
r_0	5	Radiative forcing baseline factor

Table D.2: Parameters of the replicator–climate–carbon coupled model (social subsystem).

Parameter	Value / Range	Description
<i>Social subsystem parameters</i>		
$f_{\max}$	2.0	Maximum risk perception response
ω	5.0	Steepness of logistic damage response
T_c	2.0, 3.0	Critical temperature thresholds (varied)
$n_{\text{categories}}$	2 (binary), 100 (spectrum)	Opinion categories
κ	Baseline 0.05, range [0.01, 0.10]	Social learning rate
$\alpha_{\max}$	Baseline 1.0, range [0.0, 2.0]	Climate damage sensitivity
δ	Baseline 1.0, range [0.0, 2.0]	Social reinforcement strength
$m_{\max}$	Baseline 1.0, range [0.0, 2.0]	Maximum mitigation utility

Table D.3: Parameters and numerical settings for the FJ–forest coupled model. Baselines are used unless a sweep over the stated range is performed.

Symbol Name	/	Baseline Range	/	Role / Description
<i>Forest-cover dynamics</i>				
$F(0)$		0.2		Initial forest cover
c		Baseline sweep [0.1, 0.5]	0.3,	Growth gain coefficient in $w(F)$
k		5		Steepness inside $w(F)$
v		Baseline sweep [0.0, 0.12]	0.01,	Linear loss rate
h		0.01 (fixed)		Social push on F via mean opinion
<i>Runaway / environmental drive</i>				
R_0		1		Baseline offset in r_F
$R_{\max}$		Baseline sweep [1.0, 2.0]	1.5,	Maximum runaway pressure amplitude
α		5		Runaway steepness
F_c		{0.2, 0.8}		Threshold scenarios (blue/orange in plots)
<i>FJ social dynamics and heterogeneity</i>				
Ψ		Baseline sweep [0.01, 0.10]	0.05,	Social timescale (interaction strength)
A		Baseline sweep [0.2, 1.0]	0.35,	Interaction length in W_{ij}
λ_0		Baseline sweep [0.2, 0.8]	0.5,	Mean susceptibility; individual λ_i are drawn heterogeneously around this value
$X_i^{(0)}$		Mode 1: 0; Mode 2: half -1 , half $+1$		Initial opinions (shuffled in mode 2)

D.4 Data Availability

All data and code supporting this study are available at:
<https://github.com/Yazdan-Babazadeh/Binary-Continuous-Social-Models>