Prediction of Abrupt Transitions in Climate Systems

by

Arshia Ebadi

A research project
presented to the University of Waterloo
in fulfillment of the
research paper requirement for the degree of
Master of Mathematics
in
Computational Mathematics

Waterloo, Ontario, Canada, 2024

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

Climate change refers to long-term shifts in temperature, weather patterns, and atmospheric conditions on Earth, primarily driven by human activities. While climate variability has naturally occurred throughout Earth's history, the accelerated pace and magnitude of changes in recent decades are unprecedented. This rapid warming influences diverse aspects of the planet, from melting polar ice caps and rising sea levels to shifts in precipitation patterns and more frequent extreme weather events. Climate change impacts ecosystems, biodiversity, and human societies, posing threats to agriculture, infrastructure, health, and global stability.

One of the main concerns of the scientific community is that unchecked human activity might push the system over a tipping point, resulting in a sudden and possibly irreversible shift in the climate. Understanding whether the climate is near a tipping point or not is a crucial piece of information that can establish a time frame in which we still have the ability to prevent this possibly irreversible shift. However, this task has proven to be difficult. Several statistical and machine learning methods have been previously tested as predictors for an upcoming tipping point in complex dynamical systems (such as the climate), but the need for a better approach is apparent.

In this paper we develop a transformer model for the task of tipping point prediction in climate systems. We train the model on statistical indicators extracted from simulated two-dimensional Ising models, and test it on data from the Climate Model Intercomparison Project (CMIP5). We show that our transformer model has better predictive ability than previous CNN-LSTM models tested on the same dataset.

Acknowledgements

I would like to thank my research supervisors Professor Bauch and Professor Anand for their support throughout this project. I am also grateful to Dr. Dylewsky for his extensive help with the Ising models and navigating the CMIP5 data. The raw training and testing datasets used in this project were taken from the data provided in Dylewsky et al. in [13]

I acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and thank the different climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Several papers in recent years have studied the impact of climate change on different aspects of human life. From water scarcity [2] to prolonged wildfires [1], the effects of climate change are becoming more tangible by the year, and can have significant economic and social consequences. Several important commodities such as rice, soybeans and potatoes have hit record high prices in recent years due to extreme weather events [14]. A projected climate-related reduction in yields for important crops such as corn can cause moderate to significant food insecurity in many areas of the world such as Africa and South Asia in the coming decades [18], [19].

The effects of climate change stretch far beyond reduced crop yields. In this age of misinformation, many people might consider climate-related risks to be of very low priority [12] and prioritize other things such as the economy without understanding that they are in fact connected. Recent papers have shown that climate change can cause severe economic problems [8]. Many financial institutions in the developed world have not fully accounted for climate change in their risk framework, and are thus vulnerable to climate risk. For example, after the recent chain of extreme weather events in Florida (including hurricanes Ian, Milton, and Helene) many insurance companies have suffered heavy losses, leaving them with no choice but to greatly hike insurance premiums for most homeowners and stop insuring certain areas altogether [20]. This has negatively affected many ordinary people. Prospective home buyers cannot get a mortgage from banks if their property is located in one of the uninsurable areas. On the other hand, many homeowners are now under additional financial stress given the higher insurance premiums. Several other states such as California and Louisiana are grappling with similar issues [20]. Even by the most conservative standards, the increase in extreme weather events can at least somewhat be

attributed to climate change [22]. Thus, the economy is not shielded from the negative impacts of climate change.

The aforementioned issues underline the importance of understanding the evolution of the climate. A natural cause for great concern is the possibility of a sudden and possibly irreversible shift in the climate, such as a sudden drop in rainfall or sudden increase in temperature, which can cause a number of issues as outlined above. Even though precisely predicting the evolution of a system might be too costly, a "big-picture" approach can still provide useful insights. Other researchers have worked on developing techniques to predict if a dynamical system will undergo an abrupt transition at some point in the future, without actually predicting the full evolution of the system [6]. Given that the climate is a dynamical system, previous papers have used bifurcation theory to propose certain statistical metrics (such as variance and autocorrelation) as early warning signals (EWS) for upcoming abrupt shifts [9]. More recently, the focus has shifted to using machine learning (ML) frameworks that have proved more accurate than the aforementioned statistical metrics [13], [6]. In this project, our goal is to further improve on the current ML models. To this end, we propose a modified transformer classifier architecture and compare it with the CNN-LSTM model used by Dylewsky et al. in [13].

Chapter 2

Background

This section covers the necessary background information that is required to understand both the methodology and purpose of this research project. We note that throughout this paper the phrases "tipping point", "abrupt transition" and "phase transition" are used somewhat interchangeably.

2.1 Dynamical Systems

A dynamical system is a framework used to describe how a point (or vector) in a given space evolves over time. We can consider the general system:

$$\vec{x'} = f(t, \vec{x}, \vec{\mu}) \tag{2.1}$$

where $\vec{x}$ is the state of the system at time t, x' is the rate of change of the system with respect to time and $\vec{\mu}$ is a parameter that can change the behavior of the system as we will discuss later. In general, t (and by extension f) can be continuous or discrete. For this research project, we want to ultimately work with real climate data which is in discrete format as it is gathered at fairly regular time intervals, so we take time to be discrete as well.

In practice it can be very difficult to find the function f that governs a system's dynamics just from periodic observations (which is the type of data that can be gathered in the real world). Even if f was known (in reality we would probably only be able to approximate it from the data), it might be very computationally demanding to compute

the future trajectory of the system. However, sometimes we can learn about the qualitative properties of a system without knowing its exact trajectory. One method of learning about the "big picture" is through the study of a system's equilibrium points and possible bifurcations.

2.1.1 Equilibrium States

In the context of dynamical systems, equilibrium states (or fixed points) are states where the system does not change over time. As a simple example, see:

$$x' = x^2 \tag{2.2}$$

where x is the state of the system (which we take to be a function of time) and x' is the rate of change of x. In this case x' is not directly a function of time so we call Equation (2.2) a homogenous system. In general, if the equation has an explicit dependence on time, it is called non-homogenous, as is the case with Equation (2.1). It is clear that if $x = 0$, then $x' = 0$. Thus, x will never change and we say that $x = 0$ is an equilibrium point of Equation (2.2).

Equilibrium points can be classified as stable or unstable, based on how the system behaves when it is slightly perturbed from the point.

Stable Equilibrium

An equilibrium point is said to be stable if, when the system is slightly perturbed (moved slightly away from the equilibrium), it tends to return to that equilibrium point over time. In other words, small disturbances do not cause the system to move away from the equilibrium; instead, the system tends to be attracted back to this point.

Unstable Equilibrium

An equilibrium point is said to be unstable if, when the system is slightly perturbed, it tends to move further away from that equilibrium point over time. In this case, small disturbances cause the system to deviate more and more from the equilibrium.

We will now examine the effect of small perturbations of the different equilibrium points in equation:

$$x' = x^2 - 4 \tag{2.3}$$

where one can easily see that $x' = 0$ if and only if $x = -2$ or $x = 2$. First we will focus on a neighborhood around the point $x = -2$. If we start with an x slightly to the right of the equilibrium point ($x > -2$), Equation (2.3) will give $x' < 0$, driving x back to the left and towards the equilibrium point. Similarly, if we start with $x < -2$, we will get $x' > 0$, once again driving our system towards the equilibrium point. Since all points in the neighborhood of $x = -2$ move closer to it as time goes on, we say that this equilibrium point is stable.

Now we shift our focus to a neighborhood around $x = 2$. For a small perturbation to the right ($x > 2$), we will have $x' > 0$ so the state will keep going to the right and get farther away from $x = 2$. Similarly, if we start with $x < 2$, we will have $x' < 0$ and the state will once again get farther away from the equilibrium as time passes. Thus, we call this an unstable equilibrium. The stable and unstable equilibrium of System (2.3) are highlighted in Figure 2.1.

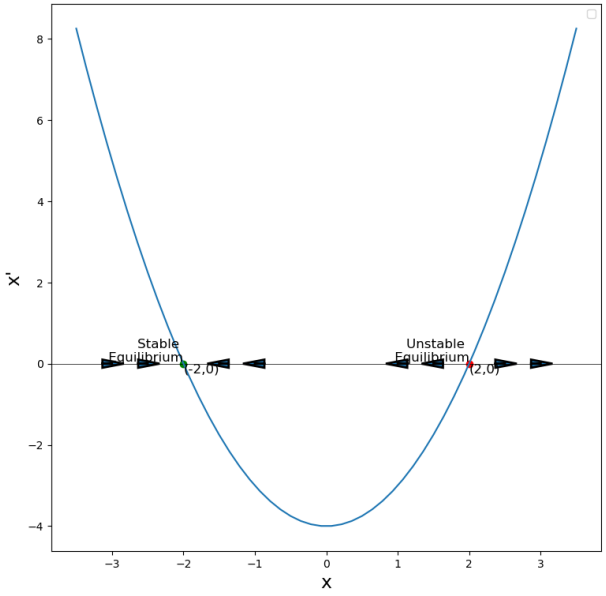

Figure 2.1: Graph of System (2.3). Arrows show trajectory of non-equilibrium states through time. $x = -2$ is a stable equilibrium and $x = 2$ is an unstable equilibrium

2.1.2 Bifurcations

Let us look at the generalized version of Equations (2.2) and (2.3) as given below:

$$x' = x^2 - \mu \tag{2.4}$$

where μ is a parameter that can take on any real value. We have already seen that different values of μ produce systems that have different types of equilibrium points. These results are summarized in Figure 2.2. For $\mu < 0$ the system will not have any equilibrium points since $x^2 - \mu$ will not have any real roots. At $\mu = 0$, there is only one unstable equilibrium point at $x = 0$ as previously discussed. Finally, for $\mu > 0$ we see a stable equilibrium at $-\sqrt{\mu}$ and an unstable equilibrium at $\sqrt{\mu}$. This qualitative change in the behavior of System (2.4) as μ flips from negative to positive called a bifurcation. Specifically, $\mu = 0$ is called the bifurcation point as it is here that equilibrium points suddenly appear (or equivalently disappear depending on your direction of approach). In this context, we call μ the order parameter.

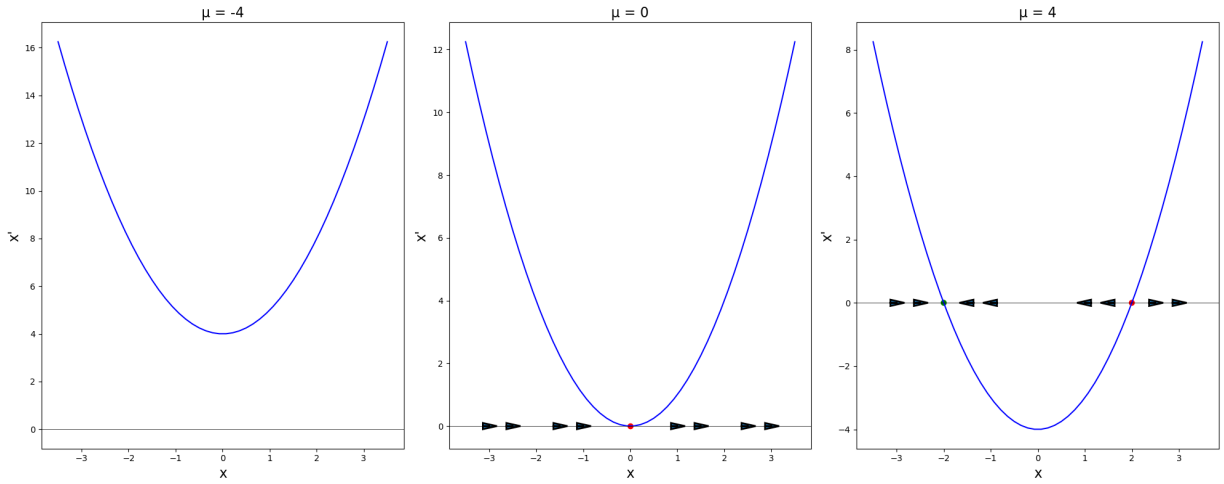

Figure 2.2: Graph of system (2.4) with equilibrium points highlighted for different values of μ.

Extending this analysis to all values of μ, we arrive at the bifurcation diagram for the saddle-node bifurcation shown in Figure 2.3, which depicts the number and position of equilibrium points as a function of the order parameter μ. This is just one of many different types of bifurcations, many of which are more complicated and some that only occur in higher-dimensional systems. In fact, higher-dimensional systems that have multiple

parameters can undergo different types of bifurcations based on which parameter is being varied. The bifurcation point for Equation (2.4) was easy to find, but for many higher-dimensional systems this task can be very difficult and sometimes not analytically feasible. For this reason, other tools have been developed to detect bifurcation points in systems, which we will discuss below.

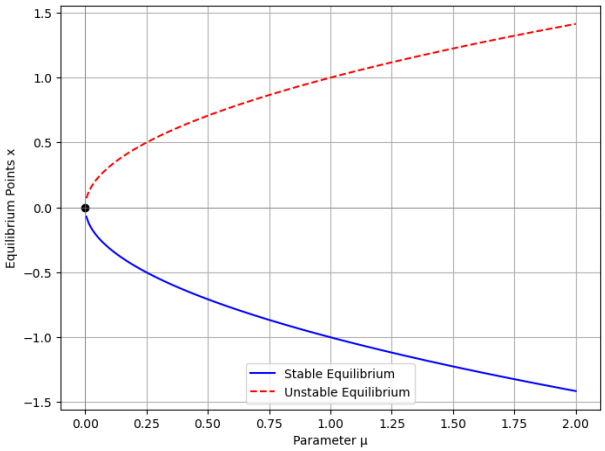

Figure 2.3: Bifurcation diagram depicting the Saddle-Node bifurcation

2.1.3 Early Warning Signals

Several tools have been developed using bifurcation theory for cases where the bifurcations cannot be easily analytically detected. Some of these methods rely on observing the behavior of the system over time and looking for certain characteristics, which have been shown to precede certain bifurcations. An example of these characteristics is the critical slowing down of a dynamical system as it gets closer to a local bifurcation [27], [23]. This is where the system becomes less resilient to perturbations as it approaches a bifurcation point (i.e. the system takes longer to recover from perturbations), and thus the variance and autocorrelation of the system see an increase. For this reason, variance and autocorrelation have previously been put to the test as early warning signals (EWS), where they have found mixed success. Some systems have been shown to exhibit EWS before going through transitions, including Earth's paleoclimate system and lake manipulation experiments [10], [7]. However, other studies have shown so-called silent phase transitions that do not exhibit early warning signals [24], [5]. This marks the need for better tools in predicting future bifurcations in dynamical systems.

2.2 Machine Learning: Transformers

Machine learning is an umbrella term encompassing different methods of constructing a model that can be fitted to observed data and subsequently used to make predictions about new (or out-of-sample) data [21]. There are also generative ML algorithms that create new data from the learned distribution but they are outside the scope of this project. In the context of ML, the prediction of future bifurcations (phase transitions) in a dynamical system can be thought of as a classification task. We can consider a set of observations (systems that we know will undergo a transition) that will be used to train an ML algorithm (in this case a transformer) to make predictions about previously unseen data (climate systems that we are interested in).

Climate data is normally presented as a time series and is inherently temporally ordered, so researchers have used different ML architectures that can specifically account for this order. One of the architectures equipped to handle temporal data is the transformer. It was originally introduced in the context of natural language processing (NLP) in [25]. It is a deep learning model that first introduced the self-attention mechanism, which allows for the model to determine the importance of an input based on other inputs that come both before and after it. Over the years, the transformer architecture has become a powerful tool for a variety of sequential tasks due to its ability to capture complex long-term dependencies in data [26].

2.3 Statistical Metrics

In this section we give a brief explanation of the different statistical metric calculations that are frequently discussed in this project.

Consider a generic data distribution where x_i is data captured at time t_i. The moments of this distribution give insight into its shape [16]. The 1st moment gives the mean or expected value of the distribution ($E[X]$), and can be calculated as follows:

$$E[X] = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.5}$$

The 2nd moment (which is conventionally calculated around the mean and is thus called 2nd central moment) gives the variance of the data and can be written as: $E[(X - E[X])^2]$ In general, the n-th central moment is given as $E[(X - E[X])^n]$. The (standardized) 3rd central moment is called the skewness, and the (standardized) 4th central moment

is called the kurtosis[16]. Although the formulas get more complicated, there are many Python packages that can calculate these moments for a given dataset.

The next statistical metric of interest is autocorrelation. The autocorrelation of a distribution is simply its correlation with a lagging copy of itself, where the amount of lag is a parameter [11]. For example the lag-2 autocorrelation of X would be its correlation with $X'' = \{x_3, x_4, x_5, ..., x_n\}$, which can be calculated as follows:

$$corr(X, X'') = \frac{cov(X, X'')}{\sigma_X \sigma_{X''}} = \frac{E[(X - E[X])(X'' - E[X''])]}{\sigma_X \sigma_{X''}} \tag{2.6}$$

where σ is the standard deviation of the distribution and can be calculated by taking the square root of the variance. Lag-α (where α is the time lag) can also be calculated using various packages in Python.

2.4 The Ising Model

The Ising model is a mathematical model used to study magnetism in materials [17]. It is a simple yet effective tool that can be used to study phase transitions. The model can be described as a collection of sites, each of which has a "spin" (also called "magnetic moment") that can be in one of two possible states: spin up (+1) or spin down (-1).

We focus on the two-dimensional Ising model, since the one-dimensional model does not go through any phase transitions, and higher-dimensional models are more complicated. In the two-dimensional model, the sites are arranged in a square lattice. Neighboring sites in the lattice interact with each other and with an external magnetic field if we choose to apply one. The formula for the Hamiltonian (total energy) of the Ising model is given below (assuming a closed system with no energy transfer from outside the system):

$$E = -\sum_{\langle i,j \rangle} J_{i,j} s_i s_j - \sum_i h_i s_i \tag{2.7}$$

where $J_{i,j}$ is the interaction energy between sites s_i and s_j (that we take to be greater than zero), h_i is the external magnetic field at site s_i (which we allow to take on any real number), and the notation $< i, j >$ means that the sites are neighbors. For this project we only considered uniform external magnetic fields, meaning that in our simulations $h_i = H$ for all i, and Equation 2.7 can be simplified to:

$$E = -\sum_{\langle i,j \rangle} J_{i,j} s_i s_j - H \sum_i s_i \tag{2.8}$$

where H is the uniform eternal magnetic field. Note that since we are talking about the two-dimensional model, each of i and j are actually themselves a 2-dimensional vector signifying the location of a site. Thus, the total energy is the sum of all interactions between neighboring sites in addition to the interaction of every site with the external field. As the model evolves through time, each site has a probability of flipping its spin based on how the flip affects the total energy of the system. If the flip causes a decrease in total energy, it will happen with probability 1. If the flip causes an increase in total energy, it will happen with a probability proportional to $e^{-\frac{1}{T}}$ (meaning that higher temperatures allow higher probability of energetically unfavorable flips).

Another property of interest in the Ising model is the net magnetization, which is simply the mean of the spins of every site in the collection. It can be easily calculated as follows:

$$M = \frac{\sum_i s_i}{n^2} \tag{2.9}$$

where n is the side length of the Ising lattice (which we take to be a square). The division by number of sites is done to keep M between -1 and 1 so that systems of different sizes can be compared. The magnetization and energy are dependent on the direction of the spins, which are themselves dependent on the external magnetic field H and temperature T. Thus, we consider H and T as parameters of the Ising model. We will now discuss how the model undergoes different phase transitions as these parameters are varied.

2.4.1 Transitions

If the temperature parameter T is kept constant and the external magnetic field H passes through its critical value H_c, a discontinuous change will appear in the energy E of the system, marking a first-order transition. This is depicted in Figure 2.4. Another way to look at this transition is through the net magnetization of the Ising lattice, which will also see a discontinuous change as shown in Figure 2.5.

The Ising model also exhibits transitions of other orders. If we keep the external magnetic field constant and vary the temperature through its critical value T_c, the heat capacity (which is the derivative of energy with respect to the temperature) will undergo a discontinuous change. This transition is considered second-order because the energy is

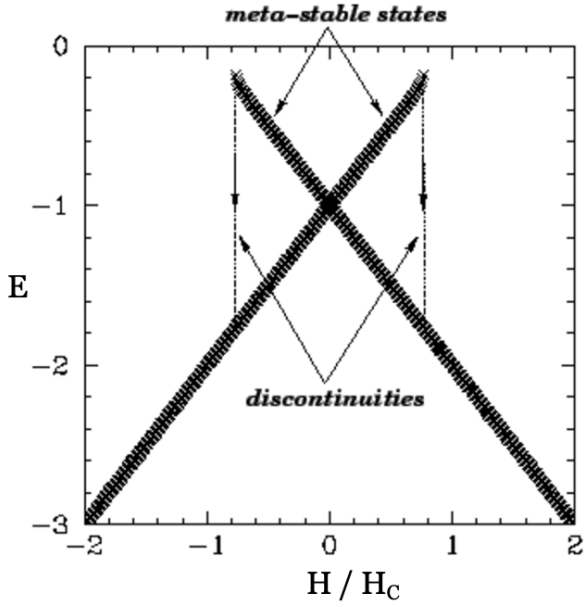

Figure 2.4: The energy (E) of an Ising lattice as a function of external magnetic field (H), at constant temperature. Discontinuous changes occur at H_C and $-H_C$ depending on if H was increasing or decreasing. Picture taken from [17]

continuous as a function of temperature and the discontinuity appears in its derivative. We will not discuss these transitions as only first-order transitions were used in this project.

2.5 Related Work and Motivation

The main motivation of this project is to develop a tool for the prediction of abrupt transitions in climate systems. As pointed out in Section 2.1, most climate models are computationally demanding to solve as they can be very high-dimensional, and produce less accurate results when we try to predict further into the future given the nature of numerical differential equation solvers. Given the specific task of tipping point detection, it is not necessary to fully solve for the state of the system in order to determine if it goes through a tipping point. Previous tools developed for this task include certain statistical metrics chosen based on results from bifurcation theory (i.e. the EWS described in Section 2.1.3) and newer ML frameworks such as the CNN-LSTM architecture proposed by Dylewsky et al. in [13]. They trained their models on EWS data obtained from Ising models, and

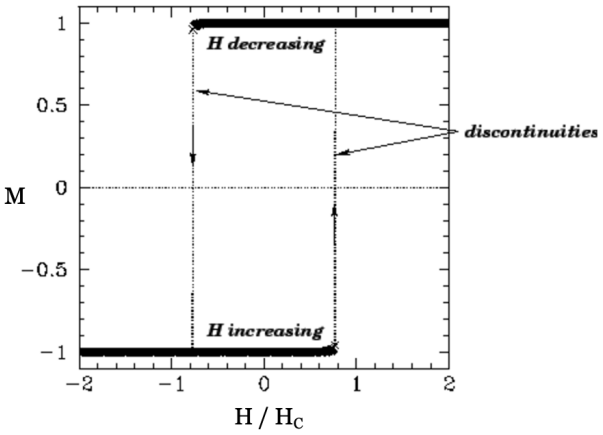

Figure 2.5: The net magnetization (M) of an Ising lattice as a function of external magnetic field (H), at constant temperature. Discontinuous changes occur at H_C and $-H_C$ depending on if H was increasing or decreasing. Picture taken from [17]

tested them on different sets of climate-related data to see if the model can accurately predict abrupt transitions. While they saw success, the goal of this paper was to construct a machine learning model that outperforms their CNN-LSTM. To this end, we propose a modified encoder-only transformer architecture and compare its performance with the CNN-LSTM. We note that they trained three different models, based on EWS in different dimensions (i.e. spatial, temporal and both), so we aimed to do the same and try to draw conclusions about which dimension contains more data regarding upcoming abrupt transitions.

Chapter 3

Methodology

This project, like most others centered around classification, consists of the following three main parts:

1. Training data
2. Classifier architecture
3. Testing Data

In this chapter we discuss each part in detail.

3.1 Training Data: Ising Model

The training data used for this project was a subset of the data created by Dylewsky et al. for training their CNN-LSTM models. They started with 128 x 128 Ising lattices and allowed them to evolve while changing the external magnetic field (H) through time. To create "null runs" they kept H away from its critical value but for "transition runs" they passed H through H_C. Importantly, the transition runs were truncated to 100 time steps before the actual transition since the whole point of the project is to predict the transitions before they actually happen, and including the transition itself in the training data would make it impossible to draw any conclusions from the models. In order to produce a certain degree of variability in the simulation results and allow for the transformers to better generalize from the data, they made several modifications to the base Ising models originally

detailed in section 2.4. Namely, the coupling coefficients ($J_{i,j}$ from Equation (2.7)) were picked from a normal distribution with randomized mean and standard deviation for each simulation. Also, random sites were "masked" (i.e. their spin was set to 0). Given that the sites only take on values of ± 1 (and 0 for masked sites), coarse-grain averaging was performed through space and time dimensions to obtain a smoother data distribution. The coarse-graining coefficients were also randomized to capture a diverse range of spatial and temporal scales. After these modifications, they ended up with 9 x 9 systems which were truncated or padded with zeros to have a length of 600 time steps, meaning that each run had a size of (600,9,9). In total, they created 2940 Ising simulations with order parameter H. This dataset was well balanced as it consisted of 1464 transition runs and 1476 null runs. The net magnetization of a transition run is shown in Figure 3.1. Note that the run itself contains the spins of the lattice sites that change over time and the net magnetization is later calculated from these spins. The magnetization is not used in the training data and is only included here to illustrate the occurrence of an abrupt transition.

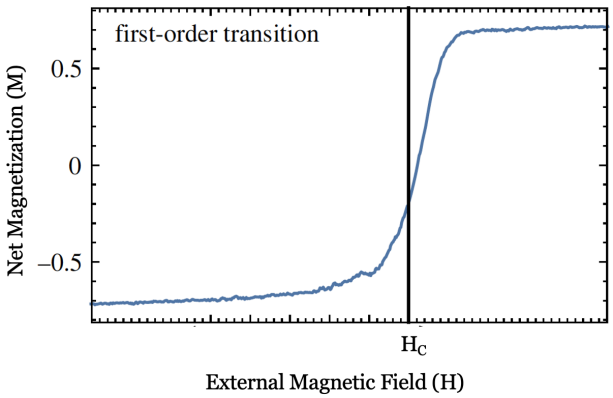

Figure 3.1: Example Ising transition run. H_C marks the inflection point of the net magnetization signifying the presence of a first-order transition.

In the next step, based on arguments detailed in Section 2.1.3, they computed the following six statistical metrics in each of the spatial and temporal domains (for a total of twelve metrics):

1. Variance
2. Skewness
3. Kurtosis

4. Lag-1 autocorrelation
5. Lag-2 autocorrelation
6. Lag-3 autocorrelation

From these metrics they produced three different data sets that we used to train three different models. The first dataset contains only the temporal statistics, so each data point is a time series with shape (600,6). The second dataset contains only the spatial statistics so it has the same shape. The third dataset contains all twelve statistics, so each data point has shape (600,12).

We trained three models as explained below:

1. Full model: Trained on both spatial and temporal statistics
2. Spatial model: Trained only on spatial statistics
3. Temporal model: Trained only on temporal statistics

3.2 Classifier Architecture: Transformer

The decision to choose a transformer is based on previous research showing that they can capture long-term dependencies in sequential data to a great extent [26]. An advantage of the transformer is the self-attention mechanism which allows it to understand the context of each time step in the time series by looking at time steps both before and after it [25]. This is in contrast to the CNN-LSTM architecture which looks at time steps one at a time and can therefore only consider the relevance of a time step with respect to the ones that came before it and not after.

The original transformer model consists of an encoder and a decoder as it was designed for language translation. The encoder is designed to extract "meaning" from the input and map it to an abstract vector space, whereas the decoder uses this "meaning" and previous inputs to generate an output sequence [25]. Our task was time series classification, so we did not want to predict future values in the time series but rather wanted to use the available time steps to come to a conclusion about the presence of an upcoming abrupt transition. In other words, we did not intend to generate the next time steps in the time series. For this reason, we decided to use an encoder-only transformer. The overall architecture is shown in Figure 3.2.

The transformer was implemented using the PyTorch library, and the hyper-parameters were selected as follows:

- batch size: 32
- Learning rate: $5 \times 10^{-}5$
- Self-attention heads: 3 per encoder layer
- Positional encoding dropout rate: 0.2 (full model), 0.25 (spatial and temporal models)
- Training epochs: 70 (full model), 125 (temporal and spatial models)

The Binary Cross-Entropy loss function and Adam optimizer were used for training the models.

3.3 Test Data: CMIP5

The test data was chosen from the CMIP5 repository (Climate Model Intercomparison Project v5), which contains a vast number of climate projections based on different assumptions regarding future human activity (i.e. anthropogenic effects). Specifically, the GFDL-ESM2M dataset from the CMIP5 repository was used to test the classifiers. This dataset contains spatiotemporal time series of different climate-related variables, summarized in Table 3.3.

4345 time series from the GFDL-ESM2M dataset were selected and processed by Dylewsky et al. to extract the same statistical metrics as outlined in Section 3.1. Abrupt transitions in the dataset were identified using the edge detection results of Bathiany et al. outlined in [3], and data was labeled accordingly. 400 random data points were then selected as the validation set to check the accuracy of the model during training.

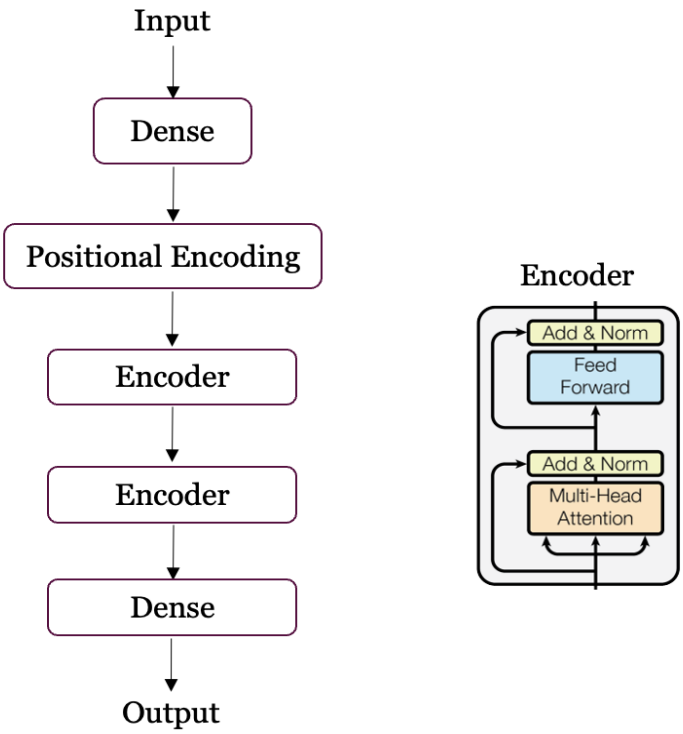

Figure 3.2: Architecture of the encoder-only transformer. The input is size (32, 600, d) where 32 is batch size and d is twelve for the full model but six for spatial and temporal models. Dense layers are also called fully-connected layers. The first dense layer has size (600, 12) for the full model but is changed to size (600, 6) for the other two. The positional encoding cell also includes a dropout component which is a tunable hyper-parameter. Encoder cells are comprised of multi-head attention and dense feed-forward layers, each followed by residual connections and normalization layers. The basic layout of an encoder layer is shown on the right (picture taken from [25] The final output (after applying the *tanh* activation function) is a scalar between zero and one which can be interpreted as the probability of an abrupt change occurring in the future.

Variable Name	Climate Realm
Near-Surface Air Temperature (height 2m)	Atmosphere
Eastward Near-Surface Wind (height 10m)	Atmosphere
Near-Surface Wind Speed (height 10m)	Atmosphere
Near-Surface Relative Humidity (height 2m)	Atmosphere
Near-Surface Specific Humidity (height 2m)	Atmosphere
Precipitation	Atmosphere
Surface Downwelling Clear-Sky Shortwave Radiation	Atmosphere
TOA Outgoing Longwave Radiation	Atmosphere
Water Vapor Path	Atmosphere
Total Cloud Fraction	Atmosphere
Condensed Water Path	Atmosphere
Ice Water Path	Atmosphere
Air Pressure at Convective Cloud Top	Atmosphere
Dissolved Inorganic Carbon Content	Ocean
Delta PCO2	Ocean
Surface Runoff	Land
Total Runoff	Land
Transpiration	Land
Leaf Area Index	Land
Surface Snow Melt	Land
Surface Upward Latent Heat Flux	Atmosphere
Surface Upward Sensible Heat Flux	Atmosphere
Fraction of Time Convection Occurs	Atmosphere
Surface Downward Latent Heat Flux	Ocean
Surface Downward Sensible Heat Flux	Ocean

Table 3.1: List of different variables studied in the GFDL-ESM2M dataset. Each datapoint in the dataset is three dimensional (latitude, longitude, and time) and contains historical and projected values of one of the listed variables.

Chapter 4

Results

The validation losses and accuracies obtained during training are shown in Figure 4.1 and Figure 4.2 respectively.

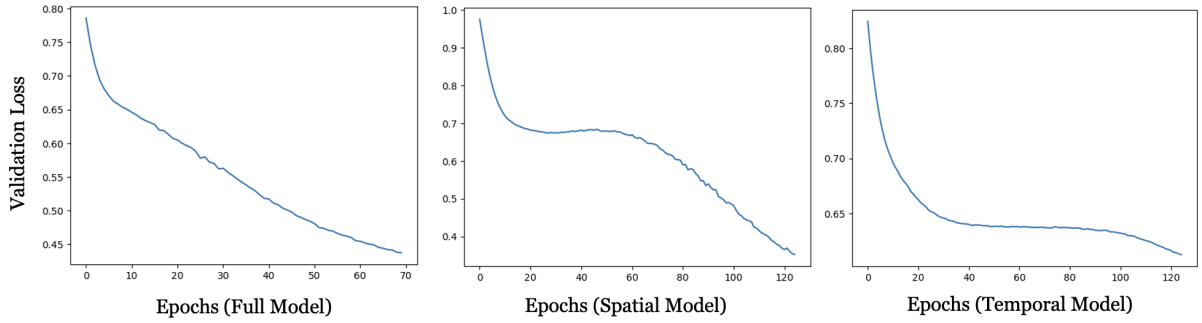

Figure 4.1: Validation loss of all three transformer models during training

A common way to assess the performance of classifier models is the use of confusion matrices. These matrices display false negatives (FN) and false positives (FP) alongside true negatives (TN) and true positives (TP). The accuracy of a model can then be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

The confusion matrices for the three transformer models along with their calculated accuracy are given in Figure 4.3. An important insight is the significantly lower accuracy of the model that was trained only on temporal statistics. This shows that the spatial

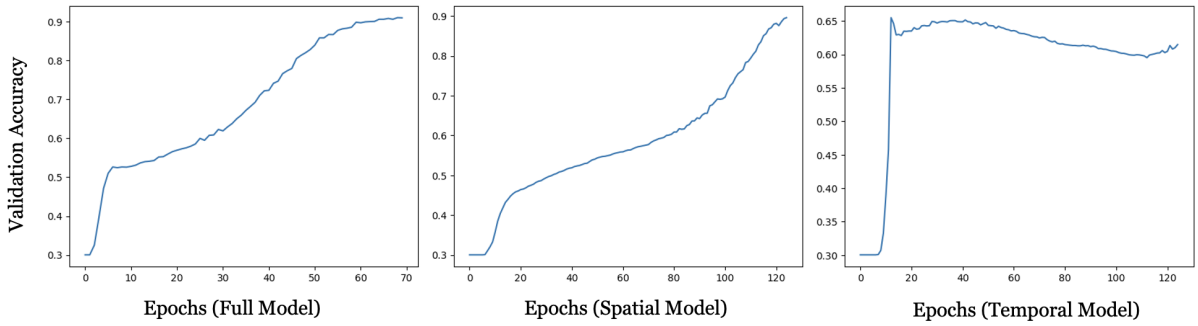

Figure 4.2: Validation accuracies of all three transformer models during training

statistics probably contain more information about the upcoming transition than the temporal statistics. Since the full model was trained on both, it should have been at least as good as the spatial model, which is in line with the results.

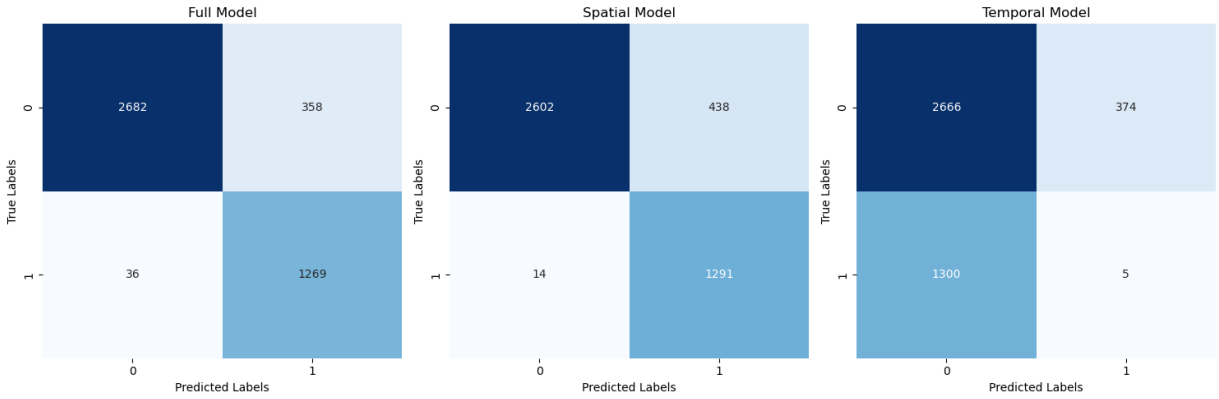

Figure 4.3: Confusion matrices for three transformer models trained on full, spatial, and temporal statistics respectively. Each matrix shows the From left to right, the accuracy of the models are 90.93%, 89.60%, and 61.47%.

No confusion matrices were provided in [13] for models tested on the GFDL-ESM2M dataset, so we cannot compare the accuracy of our models. In order to provide a direct comparison, we used another method of gauging the predictive power of a classification model, which is the receiver operating characteristic (ROC) curve. This curve is obtained by calculating the true positive rate and false positive rate at each possible discrimination threshold, and then plotting the true positive rate over the false positive rate. The area under the ROC curve (AUC) can then be used to assess the models, where a higher

AUC implies the model produces less false positives. The main point of this project was to compare the performance of the transformer models with the CNN-LSTM models developed in [13], so the ROC and AUC comparisons are shown in Figure 4.4. We see that our transformer models trained on full and spatial statistics significantly outperform their respective CNN-LSTM counterparts. The temporal model under-performs but based on its accuracy relative to the other two models, it probably does not contain much relevant information about the abrupt transitions. We did not investigate this further as any future applications of this architecture will focus on the spatial model due to its better performance.

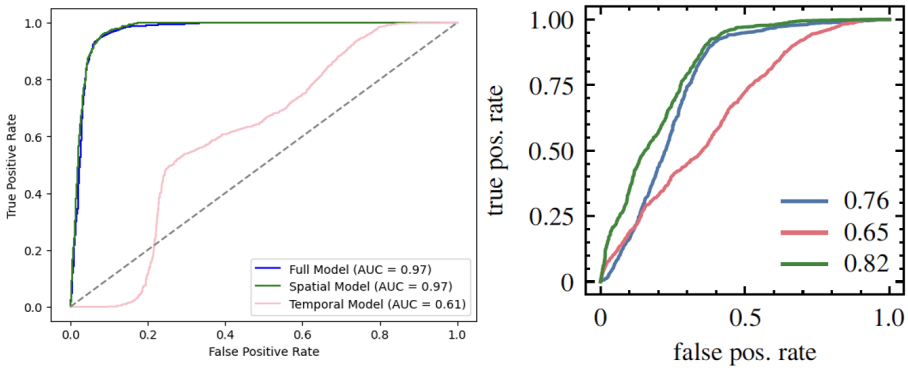

Figure 4.4: ROC curves and AUC for the three transformer models (on the left) compared to the three CNN-LSTM models developed by Dylewsky et al. (on the right). For both graphs, the blue, green and pink curves represent the full, spatial and temporal models respectively. An AUC of 0.5 means the classifier is no better than random guessing. An AUC of 1 means the classifier is perfect in that it produces no false positives.

Chapter 5

Discussion and Future Work

The ability to accurately predict upcoming abrupt transitions in climate systems can be a valuable tool as a sort of "canary in the coal mine" early warning indicator. If policymakers are made aware of the possibility of such sudden (and possibly irreversible) shifts, they will have the option to implement policies that can prevent further changes in the climate or at least mitigate damages from these changes. The transformer models developed in this project are not perfect, but they can be considered the next step in the process of developing tools to help study climate change. Just as the CNN-LSTM developed by Dylewsky et al. proved to be better than the statistical metrics previously used as EWS, this transformer architecture can be seen as an improvement on their model. Even though the current transformer models reach an accuracy of around 90% (achieved by the full model), there is still room for improvement in this regard. A logical next step would be to develop models that can predict abrupt transitions with higher accuracy. This can be done by further optimizing the hyper-parameters of the model or developing a new architecture altogether. Another shortcoming of the current model is the fact that it cannot predict exactly when an abrupt shift will occur. New models can improve on this by trying to predict how far out the possible transition will be. This task can be difficult, as it would require a multi-output model that performs both classification (to determine if a transition will occur) and regression (to give an expected time remaining before reaching the bifurcation). Nevertheless, there are many avenues for advancements in this area.

Data Availability

The model weights and training code along with the training and test datasets for all three models are available at https://github.com/Arshiaeb/Research/tree/main/Results.

References

[1] Nerilie J Abram, Benjamin J Henley, Alex Sen Gupta, Tanya JR Lippmann, Hamish Clarke, Andrew J Dowdy, Jason J Sharples, Rachael H Nolan, Tianran Zhang, Martin J Wooster, et al. Connections of climate change and variability to large and extreme forest fires in southeast Australia. *Communications Earth & Environment*, 2(1):1–17, 2021.

[2] Zunaira Asif, Zhi Chen, Rehan Sadiq, and Yinying Zhu. Climate change impacts on water resources and sustainable water management strategies in North America. *Water Resources Management*, 37(6):2771–2786, 2023.

[3] Sebastian Bathiany, Johan Hidding, and Marten Scheffer. Edge detection reveals abrupt and extreme climate events. *Journal of Climate*, 33(15):6399–6421, 2020.

[4] Austin Becker, Adolf KY Ng, Darryn McEvoy, and Jane Mullett. Implications of climate change for shipping: Ports and supply chains. *Wiley Interdisciplinary Reviews: Climate Change*, 9(2):e508, 2018.

[5] Maarten C Boerlijst, Thomas Oudman, and André M de Roos. Catastrophic collapse can occur without early warning: examples of silent catastrophes in structured ecological models. *PloS one*, 8(4):e62033, 2013.

[6] 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.

[7] Vince L Butitta, Stephen R Carpenter, Luke C Loken, Michael L Pace, and Emily H Stanley. Spatial early warning signals in a lake manipulation. *Ecosphere*, 8(10):e01941, 2017.

[8] Raffaella Calabrese, Timothy Dombrowski, Antoine Mandel, R Kelley Pace, and Luca Zanin. Impacts of extreme weather events on mortgage risks and their evolution under climate change: A case study on Florida. *European Journal of Operational Research*, 314(1):377–392, 2024.

[9] Vasilis Dakos, Stephen R Carpenter, Egbert H van Nes, and Marten Scheffer. Resilience indicators: prospects and limitations for early warnings of regime shifts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1659):20130263, 2015.

[10] 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.

[11] Cornelius Frank Dietrich. *Uncertainty, calibration and probability: the statistics of scientific and industrial measurement*. Routledge, 2017.

[12] Riley E Dunlap and Robert J Brulle. Sources and amplifiers of climate change denial. *Research handbook on communicating climate change*, pages 49–61, 2020.

[13] 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.

[14] Charlotte Edmond and Rebecca Geldard. Extreme weather is driving food prices higher. https://www.weforum.org/stories/2024/02/climate-change-food-prices-drought/, 2024.

[15] Merve Er Kara, Abhijeet Ghadge, and Umit Sezer Bititci. Modelling the impact of climate change risk on supply chain performance. *International Journal of Production Research*, 59(24):7317–7335, 2021.

[16] William Feller. *An introduction to probability theory and its applications, Volume 2*, volume 81. John Wiley & Sons, 1991.

[17] Richard Fitzpatrick and Rebecca Geldard. The Ising model. https://farside.ph.utexas.edu/teaching/329/lectures/node110.html, 2006.

[18] Tomoko Hasegawa, Gen Sakurai, Shinichiro Fujimori, Kiyoshi Takahashi, Yasuaki Hijioka, and Toshihiko Masui. Extreme climate events increase risk of global food insecurity and adaptation needs. *Nature Food*, 2(8):587–595, 2021.

[19] Jonas Jägermeyr, Christoph Müller, Alex C Ruane, Joshua Elliott, Juraj Balkovic, Oscar Castillo, Babacar Faye, Ian Foster, Christian Folberth, James A Franke, et al. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nature Food*, 2(11):873–885, 2021.

[20] Mark Nevitt and Michael Pappas. Climate risk, insurance retreat, and state response. *Ga. L. Rev.*, 58:1603, 2023.

[21] D Patil, NL Rane, P Desai, and J Rane. Machine learning and deep learning: Methods, techniques, applications, challenges, and future research opportunities. *Trustworthy Artificial Intelligence in Industry and Society*, pages 28–81, 2024.

[22] Peter Stott. How climate change affects extreme weather events. *Science*, 352(6293):1517–1518, 2016.

[23] Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC press, 2018.

[24] Bregje van der Bolt, Egbert H van Nes, and Marten Scheffer. No warning for slow transitions. *Journal of The Royal Society Interface*, 18(176):20200935, 2021.

[25] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[26] Qingsong Wen, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. Transformers in time series: A survey. *arXiv preprint arXiv:2202.07125*, 2022.

[27] C Wissel. A universal law of the characteristic return time near thresholds. *Oecologia*, 65:101–107, 1984.