Optimal Neural Network Feature Selection for Spatial-Temporal Forecasting

Eurico Covas, Emmanouil Benetos

***Abstract*—We show empirical evidence on how to construct the optimal feature selection or input representation used by the input layer of a feedforward neural network for the propose of forecasting spatial-temporal signals. The approach is based on results from dynamical systems theory, namely the non-linear embedding theorems. We demonstrate it for a variety of spatial-temporal signals, with one spatial and one temporal dimensions, and show that the optimal input layer representation consists of a grid, with spatial/temporal lags determined by the minimum of the mutual information of the spatial/temporal signals and the number of points taken in space/time decided by the embedding dimension of the signal. We present evidence of this conjecture by running a Monte Carlo simulation of several combinations of input layer feature designs and show that the one predicted by the non-linear embedding theorems seems to be optimal or close of optimal. In total we show evidence in four unrelated systems: a series of coupled Hénon maps; a series of couple Ordinary Differential Equations (Lorenz-96) phenomenologically modelling atmospheric dynamics; the Kuramoto-Sivashinsky equation, a partial differential equation used in studies of instabilities in laminar flame fronts and finally real physical data from sunspot areas in the Sun (in latitude and time) from 1874 to 2015.**

***Index Terms*—Neural network, Feedforward neural network, Input variables, Time series analysis, Forecasting, Prediction methods, Prediction algorithm, Nonlinear system, Chaos, Spatiotemporal phenomena**

I. Introduction

GIVEN a physical data set, one of the most important questions one can pose is: "Can we predict the future ?" This question can be put forward irrespectively of the fact that we may already have some insight or even be certain on what the exact model behind some or all the observed variables is. For example, for chaotic dynamical systems [1], [2], we may even have the underlying dynamics but still find it hard to predict the future, given that chaotic systems have exponential sensitivity to initial conditions. The more chaotic a system is (as measured by the positiveness of their largest Lyapunov exponents [3], [4]) the harder it gets to predict the future, even within very short time horizons. In the limit case of a random system, it is not possible to predict the future at all, although one can opine on certain future statistics [5]. For the case of weakly chaotic systems, there is an extensive literature on forecasting methods ranging from linear approximations [6]; truncated functional expansion series [7], [8]; non-linear embeddings [9]; auto-regression methods [10]; hidden Markov models [11] to state-of-the-art neural networks and deep learning methodologies [12] and many others, to long to list here.

Most literature on forecasting chaotic signals is dedicated to a single time series, or treat a collection of related time series as a non-extended set, i.e. a multivariate set of discrete variables as opposed to a spatially continuous series. For forecasting spatial-temporal chaos we refer the reader to [13]–[23] and references therein. Even rarer are attempts to forecast spatial-temporal chaos using neural networks and deep learning methodologies [24]–[34], although this field of research is clearly growing at the moment[1]. Nonetheless, this area of research is of importance, as most physical systems are spatially extended, e.g. the atmospheric system driving the Earth's weather [39]; the solar dynamo driving the Sun's sunspots [40]; and the influence of sunspots on the Earth's magnetic field via the solar wind, coronal mass ejections and solar flares – the so-called space weather [41]–[54], which may have real economical implications [55]. Nonetheless its importance, forecasting spatial-temporal chaos is difficult. The reasons are many, but mainly: first, the geometric dimension of the attractor [56] – usually quite large, the so-called curse of dimensionality [57]; and second how to choose the variables to use for forecasting, i.e., is there enough information on the same point back in time to derive the future of that particular point, or do spatial correlations and spatial propagation affect it in a way that one must take into account some spatial and temporal neighbours set to forecast the future. If this is the case, can that set of points be defined and how can it be constructed? It is this last question that we investigate in this article, in the particular context of spatial-temporal forecasting using neural networks.

Feature extraction and the design of the input representation of the input layer for a neural network is considered to be an art form, relying mostly on trial and error and domain knowledge (see [58] for examples and references). For forecasting of time series, a simple approach consists of designing the input layer as a vector of previous data using a time delay, the time delay neural network method [59]–[65]. For spatial-temporal series, one can generalize it to include temporal and spatial delays [22], [24]. This is where the connection to dynamical systems can be useful. In 1981, Takens established the theoretical

E. Covas is with the CITEUC, Geophysical and Astronomical Observatory, University of Coimbra, 3040-004, Coimbra, Portugal, and Queen Mary University London, 10 Godward Square, Mile End Rd, London E1 4FZ, e-mail: (eurico.covas@mail.com)

E. Benetos is with Queen Mary University London, 10 Godward Square, Mile End Rd, London E1 4FZ, e-mail: (emmanouil.benetos@qmul.ac.uk)

Manuscript received x, 2018; revised xx, 2018.

[1]There is also a new emerging field of research on solving PDEs (therefore implicitly predicting a spatial-temporal evolution) using deep learning – see [35]–[37] and references therein. Furthermore, notice that in this article we are concerned with the full space-time prediction, as opposed to the ongoing research on pattern recognition in moving images (2D and 3D), which attempt to pick particular features (e.g. car, pedestrian, bicycle, person, etc.) and to forecast where those features will be in subsequent images within a particular moving sequence – see [38] and reference therein.

background [66] in his embedding theorem for a mathematical method to reconstruct the dynamics of the underlying attractor of a chaotic dynamical system from a time ordered sequence of data observations. Notice the reconstruction conserves the properties of the original dynamical system up to a diffeomorphism. Further developments established a series of theorems [67]–[69] that provided the basis for a non-linear embedding and forecasting on the original variables. The theorems and related articles propose to use a time delay approach with the time lag based on the first minima of the *mutual information*[2] – see [70]–[72] – and to choose the number of points to include using the method of *false nearest neighbours* detection suggested by [73] and reported in detail in [71], [74]–[76]. Using this theoretical framework, we conjecture that this non-linear embedding method can be used to indicate what the best way to construct the feature representation for the input layer of a neural network used for forecasting both in space and time. In order to support this conjecture we, in this article, show empirical evidence for an optimal feature selection for four particular cases of a two-dimensional spatial-temporal data series s_m^n, where by two-dimensional we mean a scalar field that can be defined by a $N \times M$ matrix with components $s_m^n \in \mathbb{R}$.

Some authors discuss the use of either the mutual information and/or embedding dimension as a constraint on feature representation [62], [64], [65], [77], [78], [78], [79], [79]–[83], [83]–[93], [93], [94], [94]–[97]. Others [98] attempted to generalize the mutual information approach to higher dimensions but do not actually connect it to the problem of spatial-temporal forecasting using neural networks. There are also authors [99] that try to use neural networks to determine the optimal embedding and time delay for the purpose of local reconstruction of states with a view to forecast (the opposite of what we try to empirically demonstrate here). Fig. 6 in [100] shows how the forecasting error for a pure time series prediction changes with the delay and the number of time delay points used as an input – they use a reinforcement learning based dimension and delay estimator to derive the best dimension and delay, but do not seem to show that it is the dynamical systems derived values that are indeed optimal for forecasting neither they show any extension to spatial-temporal signals as we demonstrate in this article. Other authors [20] try to use Support Vector Machines (SVMs) to forecast spatial-temporal signals and use delays and embedding approaches to define the states vectors. In fact, Parlitz and Merkwirth [17] in their ESANN (European Symposium on Artificial Neural Networks) 2000 meeting proceedings' article mention that local reconstruction of states "...may also serve as a starting point for deriving local mathematical models in terms of polynomials, radial basis functions or neural networks...". Here we attempt to show empirical evidence that this is not just a starting point, but the optimal neural network input feature selection. We also emphasise that, as far as we are aware, all the references on neural network forecasting of spatial-temporal dynamics that use the embedding theorems and the related mutual information and the false nearest neighbours methods seem not to justify its use, i.e., the approach is explained, even suggested to be optimal but neither proven theoretically or empirically. Here we attempt to provide an empirical evidence for this optimality.

The article is divided as follows, in section II we explain our forecasting model, in section III we propose our conjecture, in section IV we show our results supporting this conjecture and finally in section V we make our concluding remarks.

Eurico Covas and Emmanouil Benetos
April 23, 2018

II. The model

The neural network architecture we choose to demonstrate our possible conjecture is a form of the basic feedforward neural network, sometimes called the time-delayed neural network [59], trained using the so-called back-propagation algorithm [101]–[104]. We focus on spatial-temporal series, so we have extended the usual time-delayed neural network to be a time and space delayed network. The overall feature representation of the network is depicted in detail in Fig. 1.

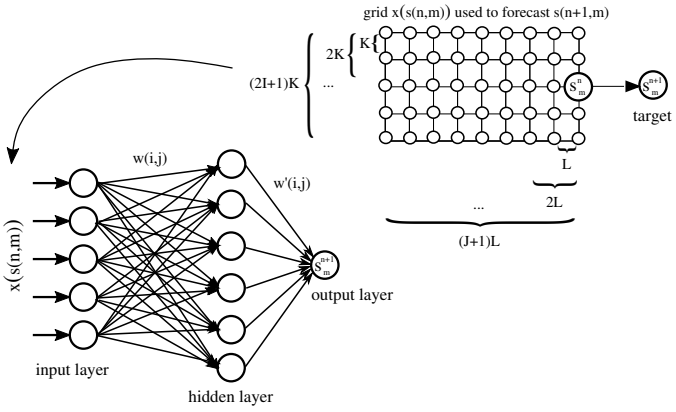

Fig. 1. Neural network architecture for forecasting spatial-temporal signals. The neural network is made of an input layer, one or more hidden layer(s) and one output layer. In this article, for simplicity, we use only one hidden layer and the output layer is made of a single neuron. Each input pattern $x(i)$ is sent to the input layer, then each of the hidden neurons values is calculated from the sum of the product of the weights by the inputs $\sum w(i,j)x(i)$ and passed via the non-linear activation function. Then the output is calculated by the product of the second set of weights times the hidden node values $\sum w'(i,j)y(i)$ again passed to another (or the same) activation function. Each input pattern $x(i)$ is actually a matrix constructed using an embedding space of spatial and temporal delays, calculated from the actual physical spatial-temporal data values $s(n,m)$. After many randomly chosen input patterns are passed via the neural network, the weights hopefully converge to an optimal training value. One can then forecast using the last time slices of the training set, and compare against the test test, the real future data set.

Under this input representation, we use the ideas proposed in [17], [18] to construct a grid of input values which are then fed to the neural network to produce a single output, the future state. Formally, let $n = 1, ..., N$ and $m = 1, ..., M$. Consider a spatial-temporal data series $\mathbf{s}$ which can be defined by a $N \times M$ matrix with components $s_m^n \in \mathbb{R}$. To these components,

[2]Notice that another non-linear dynamics technique exists to calculate this time delay, the zero of the autocorrelation function [70], [71], but essentially these two approaches are after the same objective, i.e. to select uncorrelated variables as much as possible for optimal reconstruction embedding. So in this article, we focus only on the first minima of the mutual information for simplicity of analysis.

we will call *states* of the spatial-temporal series. Consider a number $2I \in \mathbb{N}$ of neighbours in space of a given s_m^n and a number $J \in \mathbb{N}$ of temporal past neighbours relative to s_m^n (see Fig. 1 for details). For each s_m^n, we define the input (feature) vector $\mathbf{x}(s_m^n)$ with components given by s_m^n, its $2I$ spatial neighbours and its J past temporal neighbours, and with K and L being the spatial and temporal lags:

$$\begin{aligned}\mathbf{x}(s_m^n) = \{&s_{m-IK}^n, ..., s_m^n, ..., s_{m+IK}^n, \\ &s_{m-IK}^{n-L}, ..., s_m^{n-L}, ..., s_{m+IK}^{n-L}, \dots \\ &\dots, s_{m-IK}^{n-JL}, ..., s_m^{n-JL}, ..., s_{m+IK}^{n-JL}\}\end{aligned} \tag{1}$$

So, the input is a $(2I+1)(J+1)$ vector $\mathbf{x}(s_m^n)$ and the target (output) to train the network is the value s_m^{n+1}. We train the network using stochastic gradient back-propagation by running a stochastic batch where we randomly sample pairs of inputs and outputs from the training set: $\mathbf{x}(s_m^n)$ and s_m^{n+1}, respectively. Then at test time we choose inputs $\mathbf{x}(s_m^n)$, such that $n = N_{train}$, N_{train} being the number of temporal slices on the training set. As for the remaining architecture, we use one hidden layer with N_h nodes. Regarding the back-propagation hyperparameters, we included an adaptive learning rate $\eta_n = \eta/(1+n/10000)$, where the hyperparameter η is the initial learning rate and η_n is the learning rate used at time step n. We included a momentum α for faster convergence. A further hyperparameter is the choice of the activation function (see [105]), we use either a ReLu (rectified linear unit) or a logistic sigmoid function depending on the test case we are working with. We also normalize the data before passing it through the neural network, in most cases we scale it in linear fashion $x \to \alpha_{nor} + x/\beta_{nor}$, and in the case of real physical data as we will see later, we scale it in logarithmic fashion it by $x \to \alpha_{nor} + {}^{\ln(1+x)}/_{\beta_{nor}}$, where x is the initial data, and α_{nor} and β_{nor} are the arbitrary shift and scaling constants, respectively. For the weight (and bias) initialization we choose random numbers with a constant distribution between $[0, 1]$ and shifted by α_{rng} and scaled by β_{rng}. The final hyperparameter is the number of epochs taken on the stochastic gradient descent which we denote by N_{steps}. All of these hyperparameters are calibrated and fixed before we do any simulations with respect to the parameters I, J, K, L, which are auto-calibrated by the above mentioned methods derived from dynamical systems theory. In this sense, I, J, K, L are not hyperparameters of the neural network. We use the standard loss function $\mathcal{L} = \left(s_m^{n+1} - \hat{s}_m^{n+1}\right)^2$ for a prediction $\hat{s}_m^{n+1}$ centered around s_m^n using as input the features vector $\mathbf{x}(s_m^n)$ of total dimension $(2I+1)(J+1)$.

III. Conjecture

Once we do a forecast, we then compare the goodness of fit by first visual inspection and second by numerically calculating the so-called structural similarity SSIM(x, y) which has been proposed by [106] and used already in the context of spatial-temporal forecasting in [22], [24]. It has also been used in the context of deep learning used for enhancing resolution on two dimensional images [107] and restoring missing data in images [108]. For details on the SSIM measure see [106], [109], [110]. The SSIM index is a metric quantity used to calculate the perceived quality of digital images and videos. It allows two images to be compared and provides a value of their similarity - a value of SSIM $= 1$ corresponds to the case of two perfectly identical images. We use it by calculating the SSIM(x, y) between the entire test set and the forecast set, since these can be interpreted as images (one spatial dimension/one temporal dimension).

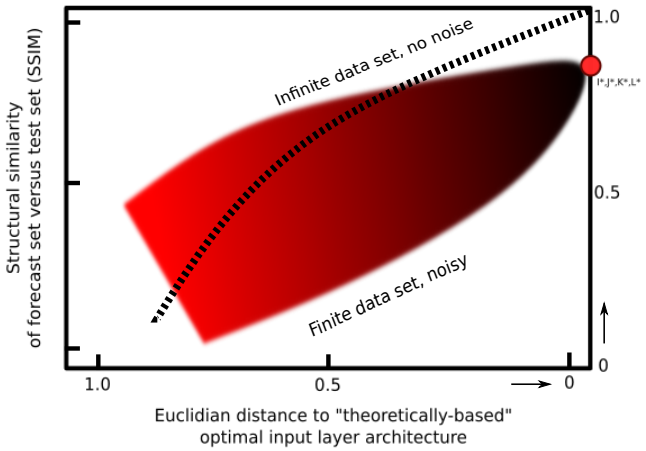

Fig. 2. Our main conjecture. For a infinite noiseless training set, the SSIM approaches SSIM $\to$ 1. For real data sets, there is a dispersion of the SSIM versus some reasonable metric constructed to represent the distance between any feature selection (e.g. $d_e\sqrt{(I-I^*)^2+(J-J^*)^2+(K-K^*)^2+(L-L^*)^2}$).

Here we propose that the optimal time delay/spatial delays (L and K, respectively) must be the ones based on the first minima of the mutual information [70]–[72] and that the optimal number of temporal/spatial points to use (J and I, respectively) must be the ones based on the the method of method of false nearest neighbours detection [71], [73]–[76]. The mutual information is calculated by taking a s^i, a one-dimensional data set, and s^{i+L}, the related L-lagged data set. Given a measurement s^i, the amount of information $I(L)$ is the number of bits on s^{i+L}, on average, that can be predicted. We then average over space and take the first minima of $\langle I(L)\rangle$, or, in the absense of a clear minima, take the L temporal lag for which the $\langle I(L)\rangle$ drops significantly and starts to plateau. This calculates L^*, the optimal time delay. Conversely, we calculate K^* by calculating the spatial lag K for which we obtain the first minima of the time-averaged mutual information $\langle I(K)\rangle$. Once the optimal spatial and temporal lags K^* and L^* are calculated, we calibrate the minimum embedding dimension, or in other words, the number of spatial and temporal neighbours in optimal phase space reconstruction. We use the method of false neighbours [73]–[75], which determines that falsely apparent close neighbours have been eliminated by virtue of projecting the full orbit in a increasing higher dimensional embedding phase space. This gives us the J^*, the optimal number of time slices to take, and I^*, the optimal number of spatial slices to take in our $\mathbf{x}(s_m^n)$ optimal reconstruction.

In this article, we conjecture that as any set of input representation "approaches" the optimal one, then SSIM $\to 1$.

In the case of finite training sets and/or noisy training sets SSIM $\rightarrow x < 1$, where x is the best forecast possible given the data set. Visually, we believe that the SSIM versus some reasonable metric constructed to represent the distance between any input representation and the optimal input representation will show a skewed bell shape as depicted in Fig. 2. In this conjecture, we use the most obvious candidate to represent the distance between any input representation and the optimal input representation, the Euclidian distance given by $d_e = \sqrt{(I-I^*)^2+(J-J^*)^2+(K-K^*)^2+(L-L^*)^2}$, where I,J,K,L are the parameters for each representation and I^*,J^*,K^*,L^* are the ones derived from the dynamical systems theory. We also verified that other reasonable metrics, in particular the Manhattan distance [111], did not change the results qualitatively.

IV. Results

In order to empirically substantiate our conjecture, we take four examples of spatial-temporal series and attempt to forecast using our feedforward neural network. First, we calculate the optimal time delay/spatial delays (L^* and K^*, respectively) using the first minima of the mutual information, then we calculate the optimal number of temporal/spatial points to use (J^* and I^*, respectively) using the method of false nearest neighbours. We then use a Monte Carlo simulation on each one of our four examples, sampling random values of the key feature selection parameters: I, J, K, L and calculate the values $d_e(I,J,K,L)$ and SSIM(I,J,K,L). We plot the latter as a function of the former to compare against our conjecture as depicted in Fig. 2.

We first take a physical system example, a real data example, and then we progress from "simpler" systems (coupled maps) capable of generating spatial-temporal chaos to more "complex" systems (coupled Ordinary Differential Equations - ODEs) to really "complex" systems (Partial Differential Equations - PDEs). This is partially motivated by results in the literature that show that general universalities are present in different levels of simplification of physical models [112], from the original PDEs to truncated ODE expansions (e.g. spectral method expansions [113]) to the most extreme simplification or discretization such as maps which capture the essence of the problem(s). In all cases we take examples with one spatial and one temporal dimension. However, we believe that our conjecture will extend to multiple spatial dimensions. Notice that here we are not trying to demonstrate that neural networks, and in particular feedforward neural networks can perform well in predicting spatial-temporal chaos (as this has been demonstrated in the literature already), but rather to show that the optimal choice of the input layer features is given by dynamical systems theory and does not need to be another neural network hyperparameter calculated by the "dark art" of trial and error.

A. Sunspot data - a physical system example

The first example we take is a physical real data example based on a previous article of one of us [24], where a neural network using the type of input representation above (Fig. 1) was used to forecast sunspot areas $A(t,\theta)$ in our Sun in both space (θ latitude) and time (Carrington Rotation index[3]). This sunspot data is usually called the "butterfly diagram" due to its butterfly wings like appearance [114]. One can see how this butterfly diagram looks like in [115].

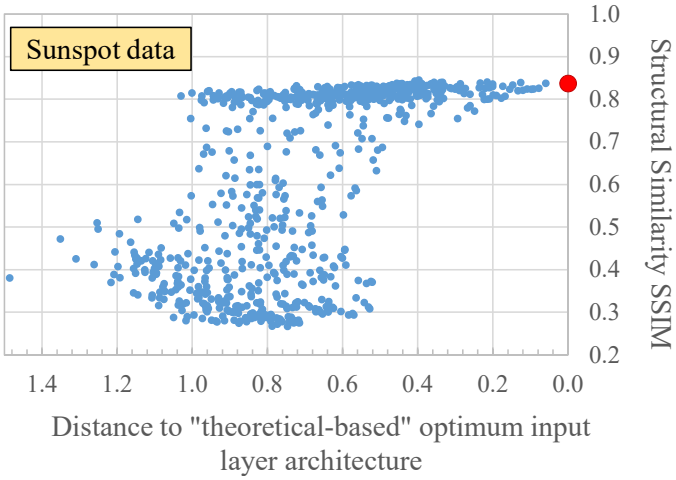

Fig. 3. Monte Carlo simulation of different input representations of the input layer for the neural network forecast for the sunspot data. It shows the structural similarity (SSIM) against how far (in an Euclidean space metric) the particular parameters of a particular run were from the supposely optimal input representation parameters (red dot).

Sunspot data is regularly seen as a benchmark for time series forecasting, given its chaotic nature and that it considered to be among the longest continuously recorded daily measurement made in science [116]. Many authors [78], [83], [88], [90], [91], [93], [94], [96], [97], [117]–[159] have already attempted to use neural networks to forecast aspects of the sunspot cycle, although as far as we are aware, none in both space and time having restricted themselves to using these neural networks to forecast mostly either the sunspot number or the sunspot areas as a function of time. There is only one example [24], as far as we are aware, of actual spatial-temporal forecasts using neural networks (see also [143], [144] where a neural network forecast of the magnetic flux, which is related to sunspots, is forecast for latitude/longitude datasets). There are also a few examples of forecasting the butterfly diagram sunspot data in both space and time (latitude/time) [22], [160]–[164] but none of these used neural networks, rather all of those used other statistical methods or numerical physical modelling.

We take as a "training set" the data from the year 1874 to approximately 1997 (i.e. the first 1646 Carrington Rotations). We then attempt to reproduce or forecast the sunspot area butterfly diagram from Carrington Rotation 1921 up to 2162 (the last one corresponding approximately to the year 2015); that is, we use 1646 time slices (≈ 122.92 years) to reproduce

[3]Given that the surface solar rotation varies with time and latitude, any approach of comparing positions on the Sun over a period of time is necessarily subjective. Therefore, solar rotation is arbitrarily taken to be 27.2752316 days for the purpose of Carrington rotations. Each solar rotation is given a number, the so called Carrington Rotation Number, starting from 9th November, 1853.

the next 242 time slices ($\approx$ 18.07 years)[4]. The training set corresponds to around 12 solar cycles (cycle 11 to 22), while the "forecasting set" equates to around 1.5 cycles (cycle 23 and half of cycle 24). The entire dataset, including the training and forecasting sets, is a grid $x_j^i = x(i,j)$, with $i = 1888$ and $j = 50$. The training set is a grid $x(1646, 50)$. For this case the optimal values were $I^* = 2$, $J^* = 6$, $K^* = 9$ and $L^* = 70$ as calculated in [22]. The hyperparameters of the neural network were: $N_h = 70$, $\eta = 0.3$, $\alpha = 0.01$, a logarithmic normalization of the inputs scaled with $\alpha_{nor} = 10$ and $\beta_{nor} = 0$, weight initialization with $\alpha_{rng} = 10^{-2}$ and $\beta_{rng} = -0.5$ and N_{steps} =1,000,000 . We used the logistic sigmoid function as the activation on both the hidden and output layers.

The Monte Carlo results are depicted in Fig. 3 showing runs with different I, J, K, L and plotting the SSIM versus the distance to the optimal input feature selection parameters (I^*,J^*,K^*,L^*) given by the dynamical systems theory. It shows a reasonable expected dispersion as conjectured and a good convergence to the highest SSIM value we could obtain for this particular slicing of the training and forecast sets SSIM = 0.836876152. From the figure, there seems to be also two clusters of behaviour, and at closer inspection, we found that the cluster with lower SSIM is basically a set of very bad forecasts, with none of the characteristics of the real sunspot behaviour (the 11 year-like cycle and the migration to the latitudinal equator), while the higher SSIM cluster corresponds to visually recognizable sunspot butterfly-like diagrams.

These results were quite satisfactory and inspired us to attempt to check the existence of an universality of behaviour across dynamical systems, by examining other unrelated synthetic generated data sets. We continue below to these attempts.

B. Coupled Hénon maps - a discrete-time dynamical system

Motivated by having a real case from a physical system, we then tried to investigate if this same conjecture holds in a very simplified example of a spatial-temporal model. Coupled maps are widely used as models of spatial-temporal chaos and pattern/structure formation [165]–[167]. Following [17], [18] we take a lattice of $M = 100$ coupled Hénon maps:

$$u_m^{n+1} = 1 - 1.45\left[\frac{1}{2}u_m^n + \frac{u_{m-1}^n + u_{m+1}^n}{4}\right]^2 + 0.3v_m^n, \quad (2)$$
$$v_m^{n+1} = u_m^n.$$

with fixed boundary conditions $u_1^n = u_M^n = \frac{1}{2}$ and $v_1^n = v_M^n = 0$. The initial values for rest of the variables $u_{m\neq 1,M}^{n=0}$ and $v_{m\neq 1,M}^{n=0}$ is taken from a random constant distribution in the range $[0, 1[$.

We run the synthetic data generation for $N = 531$ time steps, and divided the set into $N_{\text{train}} = 500$ time steps for the training set and $N_{\text{test}} = 31$ time steps for the test set. The other parameters of the neural network were: $N_h = 10$, $\eta = 0.1$, $\alpha = 0$, a linear input normalization scaling with $\alpha_{nor} = 2.947992$, $\beta_{nor} = 0.515$, $\alpha_{rng} = 10^{-3}$, $\beta_{rng} = -0.5$

[4]We use exactly the same training set/forecast set slicing as in [22], [24] for consistency, even if more data is already available at this time.

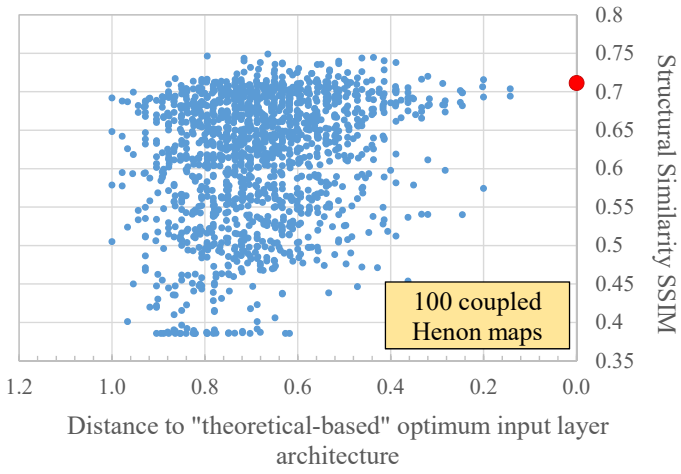

Fig. 4. Monte Carlo simulation of different input representations of the input layer for the neural network forecast for a series of 100 coupled Hénon maps. It shows the structural similarity (SSIM) against how far (in a Euclidean space metric) the particular parameters of a particular run was from the supposedly optimal input representation parameters (red dot). The green line (trendline) seems to show that as the parameters of a randomly chosen input representation get close to the supposedly optimal input representation ones, the SSIM converges to what seems to be the best possible forecast value given the limited dataset.

and N_{steps} =1,000,000 . We used the ReLu function as the activation on both the hidden and output layers.

For this case the optimal values given by the mutual information and the false neighbours methods were $I^* = 1$, $J^* = 3$, $K^* = 2$ and $L^* = 3$. The results of the Monte Carlo simulation for different I, J, K and L are depicted in Fig. 3. It again shows a dispersion as conjectured and a reasonable convergence to the highest SSIM value we could obtain for this particular slicing of the training and forecast sets SSIM = 0.71139101.

Results seems to suggest the same structure as depicted in our conjecture diagram and in the previous results for sunspots. We now move below to a more complex model, a coupled set of ODEs.

C. Coupled Ordinary Differential Equations - Lorenz-96 model

For the spatially extended coupled ODEs model we used a well-known 40-coupled ODE dynamical system proposed by Edward Lorenz in 1996 [168]:

$$\frac{dx_j}{dt} = (x_{j+1} - x_{j-2})\,x_{j-1} - x_j + F,\ j = 1, \ldots, N = 40, \quad (3)$$

where $x_{-1} = x_{N-1}$, $x_0 = x_N$ and $x_{N+1} = x_1$ and F is a forcing term. We use the forcing $F = 5$ to get some interesting behaviour in space and time. We used a time step $\Delta t = 0.05$ and we have integrated this equation using J. Amezcua's Matlab code as given in [169]. It uses the Runge-Kutta 4-step method.

We run the synthetic data generation for $N = 531$ time steps, and divided the set into $N_{\text{train}} = 500$ time steps for the training set and $N_{\text{test}} = 31$ time steps for the test set. The other parameters of the neural network were: $N_h = 10$,

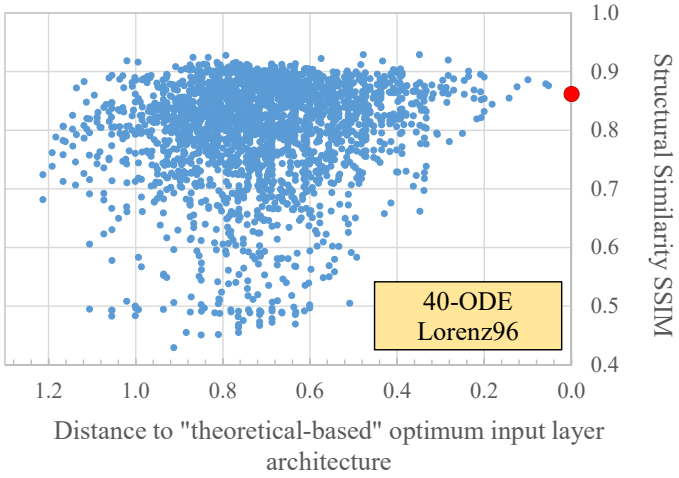

Fig. 5. Monte Carlo simulation of different input representations of the input layer for the neural network forecast for the 40-ODE Lorenz 96 system. It shows the structural similarity (SSIM) against how far (in a Euclidean space metric) the particular parameters of a particular run was from the supposedly optimal input representation parameters (red dot). The green line (trendline) seems to show that as the parameters of a randomly chosen input representation get close to the supposedly optimal input representation ones, the SSIM converges to what seems to be the best possible forecast value given the limited (and noisy) dataset.

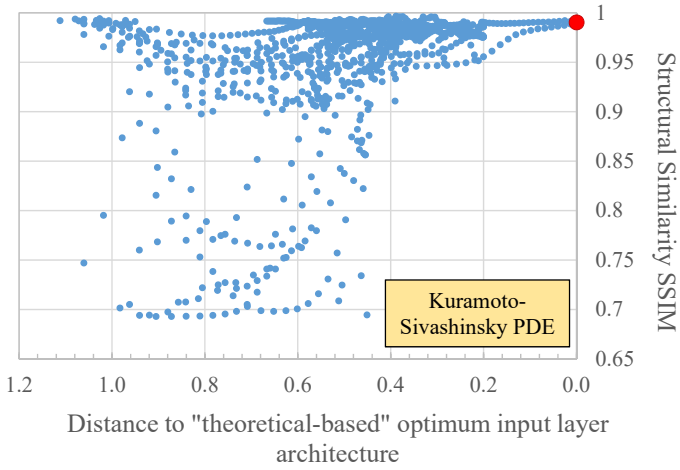

Fig. 6. Monte Carlo simulation of different input representations of the input layer for the neural network forecast for Kuramoto-Sivashinsky with $L = 22$ system. It shows the structural similarity (SSIM) against how far (in a Euclidean space metric) the particular parameters of a particular run was from the supposedly optimal input representation parameters (red dot). The green line (trendline) seems to show that as the parameters of a randomly chosen input representation get close to the supposedly optimal input representation ones, the SSIM converges to what seems to be the best possible forecast value given the limited (and noisy) dataset.

$\eta = 0.05$, $\alpha = 0.001$, a linear normalization input scaling with $\alpha_{nor} = 10$ and $\beta_{nor} = 0.430$, weight initialization with $\alpha_{rng} = 10^{-3}$ and $\beta_{rng} = -0.5$ and $N_{\text{steps}} = 100,000$. We used the ReLU function as the activation on both the hidden and output layers.

For this case the optimal values obtained before the Monte Carlo simulation from the mutual information and false neighbours methods were $I^* = 2$, $J^* = 2$, $K^* = 1$ and $L^* = 9$. The results of the random sampling of I, J, K, L in the simulation are depicted in Fig. 5. It shows a dispersion as conjectured and a quite a good convergence to the highest SSIM value we could obtain for this particular slicing of the training and forecast sets: SSIM $= 0.861844038$. Results seems to suggest the same structure as depicted in our conjecture diagram and in the previous results for sunspots and the coupled Hénon maps.

D. Partial Differential Equations - Kuramoto-Sivashinsky model

Finally we take a full PDE system, the Kuramoto-Sivashinsky model [170], [171], a very well-known system capable of spatiatemporal chaos and complex spatial-temporal dynamics. It is a fourth-order nonlinear PDE introduced in the 1970s by Yoshiki Kuramoto and Gregory Sivashinsky to model the diffusive instabilities in a laminar flame front. The model is described by the following equation:

$$\frac{\partial u(x,t)}{\partial t} = -\frac{\partial^4 u(x,t)}{\partial x^4} - \frac{\partial^2 u(x,t)}{\partial x^2} - u(x,t)\frac{\partial u(x,t)}{\partial x}, \quad (4)$$

where $x \in [-\frac{L}{2}, +\frac{L}{2}]$ with a period boundary condition $u(x + L, t) = u(x, t)$. The nature of solutions depends on the system size L and on the initial $u(x, t = 0)$. We have integrated this equation by taking an exponential time difference Runge-Kutta 4th order method (ETDRK4) using the Matlab code by P. Cvitanović as given in [172] and taking a time step of $\Delta t = 0.5$, $L = 22$ fourier modes which are known to produce a "turbulent" or chaotic behaviour and a initial condition $u(x, t = 0) = 10^{-5}$ for $x \in [5, 15]$, the remain being $u(x, t = 0) = 0$.

We run the simulation for $N = 531$ time steps, and divided the set into $N_{\text{train}} = 500$ time steps for the training set and $N_{\text{test}} = 31$ time steps for the test set. The other parameters of the neural network were: $N_h = 50$, $\eta = 0.1$, $\alpha = 0$, a linear normalization input scaling with $\alpha_{nor} = 5.8472$ and $\beta_{nor} = 0.5$, weight initialization with $\alpha_{rng} = 10^{-3}$ and $\beta_{rng} = -0.5$ and N_{steps} =1,000,000. We used the ReLU function as the activation on both the hidden and output layers.

The results of the Monte Carlo simulation can be seen in Fig. 6. For this case the optimal values obtained before we run the Monte Carlo simulation were $I^* = 1$, $J^* = 2$, $K^* = 2$ and $L^* = 39$. It again shows a dispersion as conjectured and a excellent convergence to the highest SSIM value we could obtain for this particular slicing of the training and test sets: a surprising high value of SSIM $= 0.990264382$. Results seems to suggest the same structure as depicted in our conjecture diagram and in the previous results for sunspots, the coupled Hénon maps and coupled ODEs.

V. Conclusion

We have shown empirical evidence for the existence of an optimal feature selection for the input layer of feedforward neural networks used to forecast spatial-temporal series. We believe that the selection of the features of the input layer can be uniquely determined by the data itself, using two techniques from dynamical systems embedding theory: the mutual information and the false neighbours methods. The former procedure determines the temporal and spatial delays

to take when selecting features, while the latter determines the number of data points in space and time to be taken as inputs. We conjecture that this optimal feature selection gives the best forecast, as measured by a standard image similarity index. We also conjecture that the shape of the dispersion of points on a Monte Carlo simulation across all possible feature selections on a plot of the similarity index versus the distance to optimal feature selection is a skewed bell shape with the highest value being the optimal feature selection/maximum similarity index.

In order to substantiate our conjecture, we choose four unrelated systems, in order of complexity: a set of spatially extended coupled maps; a set of spatially extended coupled ODEs; a one-dimensional spatial PDE and a real spatial-temporal data set from sunspots areas in our Sun. In all four cases, we were able to first use the mutual information and the false neighbours methods to determine the four parameters defining the input layer feature selection[5]. After calibration of the hyperparameters we then were able to forecast reasonably the test set, although this is not the objective or primary goal of this article. We then show that for a random Monte Carlo simulation across possible feature selections, the neural network did not, as expected, forecast as well as it did for the specific set of optimal four parameters given by dynamical systems theory. As conjectured, the Monte Carlo simulations show that the shape of the distribution of points was a skewed bell shape with the highest value being the optimal feature selection/maximum similarity index (subject to minor variations due to noise and the finiteness of the dataset).

Given how important spatial-temporal systems are and how we want to forecast the future as accurately as possible it is quite important to attempt to reduce the number of hyperparameters in neural network prediction, and to try to constraint the feature selection from the data properties only. If indeed our conjecture turns out to be true, it would remove the input layer feature selection as another free parameter in the already complex process of choosing the details of the neural network to use for forecasting.

In this article we have focused first and foremost in establishing empirical evidence for our conjecture, within a simple framework of feedforward neural networks with one hidden layer for the purpose of prediction in one spatial and one temporal dimensions. Naturally, there are many clear extensions to our research. First to use deeper networks with more hidden layers to possible tackle systems which are hyperchaotic (i.e. with multiple positive Lyapunov exponents). Second, to attempt to extend the conjecture with empirical evidence in high dimensions, e.g. 3+1 dimensional weather systems. Third, to extend the conjecture to more complex and fashionable neural networks models, such as recurrent neural networks, particularly echo state networks and long short-term memory networks. Fourth and last but not least, to prove the conjecture would show how dynamical systems theory can clarify the so call "dark arts" in neural network feature construction. These objectives are however, outside the scope of this research article and will be pursued later.

[5]We have four parameters for the feature selection in these cases, with one temporal and one spatial dimension. For higher dimensional systems, there will be more parameters, the exact number being double the number of dimensions of the system.

ACKNOWLEDGMENT

We would like to thank Prof. Reza Tavakol from Queen Mary University London for very useful discussions on forecasting. We also thank Dr. David Hathaway from NASA's Ames Research Centre for providing the sunspot data on which some of the results in this article are based upon. CITEUC is funded by National Funds through FCT - Foundation for Science and Technology (project: UID/Multi/00611/2013) and FEDER - European Regional Development Fund through COMPETE 2020 - Operational Programme Competitiveness and Internationalization (project: POCI-01-0145-FEDER-006922).

REFERENCES

[1] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems, 2nd Edition*. CRC Press, 2003.

[2] P. Manneville, *Instabilities, Chaos and Turbulence: AN Introduction to Nonlinear Dynamics and Complex Systems*. World Scientific Press, Oct. 2004.

[3] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D Nonlinear Phenomena*, vol. 16, pp. 285–317, Jul. 1985.

[4] H. Kantz, "A robust method to estimate the maximal Lyapunov exponent of a time series," *Physics Letters A*, vol. 185, pp. 77–87, Jan. 1994.

[5] I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes (Dover Books on Mathematics)*. Dover Publications, 1996.

[6] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, April 1975.

[7] M. J. D. Powell, "Radial basis functions for multivariable interpolation: A review," in *Algorithms for Approximation*, J. C. Mason and M. G. Cox, Eds. New York, NY, USA: Clarendon Press, 1987, pp. 143–167.

[8] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, vol. 2, 1988.

[9] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Phys. Rev. Lett.*, vol. 59, pp. 845–848, Aug 1987.

[10] G. Box, G. M. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting & Control (3rd Edition)*. Prentice Hall, 1994.

[11] L. Rabiner and B. Juang, "An introduction to hidden markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, Jan 1986.

[12] M. Lngkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters*, vol. 42, pp. 11 – 24, 2014.

[13] D. M. Rubin, "Use of forecasting signatures to help distinguish periodicity, randomness, and chaos in ripples and other spatial patterns," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 2, no. 4, pp. 525–535, 1992.

[14] U. Parlitz and G. Mayer-Kress, "Predicting low-dimensional spatiotemporal dynamics using discrete wavelet transforms," *Phys. Rev. E*, vol. 51, pp. R2709–R2711, Apr 1995.

[15] C. López, A. Álvarez, and E. Hernández-García, "Forecasting confined spatiotemporal chaos with genetic algorithms," *Phys. Rev. Lett.*, vol. 85, pp. 2300–2303, Sep 2000.

[16] S. rstavik and J. Stark, "Reconstruction and cross-prediction in coupled map lattices using spatio-temporal embedding techniques," *Physics Letters A*, vol. 247, no. 1, pp. 145 – 160, 1998.

[17] U. Parlitz and C. Merkwirth, "Nonlinear prediction of spatio-temporal time series," in *ESANN*, 2000.

[18] U. Parlitz and C. Merkwirth, "Prediction of Spatiotemporal Time Series Based on Reconstructed Local States," *Physical Review Letters*, vol. 84, pp. 1890–1893, Feb. 2000.

[19] E. Covas and F. Mena, "Forecasting of yield curves using local state space reconstruction," in *Dynamics, Games and Science I*, ser. Springer Proceedings in Mathematics, M. M. Peixoto, A. A. Pinto, and D. A. Rand, Eds. Springer Berlin Heidelberg, 2011, vol. 1, pp. 243–251.

[20] Y. Xia, H. Leung, and H. Chan, "A prediction fusion method for reconstructing spatial temporal dynamics using support vector machines," *IEEE Trans. on Circuits and Systems*, vol. 53-II, pp. 62–66, 2006.

[21] D. Gladish and C. Wikle, "Physically motivated scale interaction parameterization in reduced rank quadratic nonlinear dynamic spatio-temporal models," *Environmetrics*, vol. 25, no. 4, pp. 230–244, 2014.

[22] E. Covas, "Spatial-temporal forecasting the sunspot diagram," *A&A*, vol. 605, p. A44, Sep. 2017.

[23] R. A. Richardson, "Sparsity in nonlinear dynamic spatiotemporal models using implied advection," *Environmetrics*, vol. 28, no. 6, pp. e2456–n/a, 2017, e2456 env.2456.

[24] E. Covas, N. Peixinho, and J. Fernandes, "Neural network forecast of the sunspot diagram," *ArXiv e-prints*, Jan. 2018.

[25] P. L. McDermott and C. K. Wikle, "An Ensemble Quadratic Echo State Network for Nonlinear Spatio-Temporal Forecasting," *ArXiv e-prints*, Aug. 2017.

[26] P. L. McDermott and C. K. Wikle, "Bayesian Recurrent Neural Network Models for Forecasting and Quantifying Uncertainty in Spatial-Temporal Data," *ArXiv e-prints*, Nov. 2017.

[27] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations," *ArXiv e-prints*, Nov. 2017.

[28] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations," *ArXiv e-prints*, Nov. 2017.

[29] Z. Long, Y. Lu, X. Ma, and B. Dong, "PDE-Net: Learning PDEs from Data," *ArXiv e-prints*, Oct. 2017.

[30] J. Cao, D. J. Farnham, and U. Lall, "Spatial-temporal wind field prediction by Artificial Neural Networks," *ArXiv e-prints*, Dec. 2017.

[31] A. Ghaderi, B. M. Sanandaji, and F. Ghaderi, "Deep forecast: Deep learning-based spatio-temporal forecasting," in *The 34th International Conference on Machine Learning (ICML), Time series Workshop*, 2017.

[32] Z. Lu, J. Pathak, B. Hunt, M. Girvan, R. Brockett, and E. Ott, "Reservoir observers: Model-free inference of unmeasured variables in chaotic systems," *Chaos*, vol. 27, no. 4, p. 041102, Apr. 2017.

[33] M. Raissi and G. E. Karniadakis, "Hidden physics models: Machine learning of nonlinear partial differential equations," *Journal of Computational Physics*, vol. 357, pp. 125–141, Mar. 2018.

[34] M. Raissi, "Deep Hidden Physics Models: Deep Learning of Nonlinear Partial Differential Equations," *ArXiv e-prints*, Jan. 2018.

[35] C. Beck, W. E, and A. Jentzen, "Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations," *ArXiv e-prints*, Sep. 2017.

[36] J. Sirignano and K. Spiliopoulos, "DGM: A deep learning algorithm for solving partial differential equations," *ArXiv e-prints*, Aug. 2017.

[37] W. E, J. Han, and A. Jentzen, "Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations," *ArXiv e-prints*, Jun. 2017.

[38] C. Li, B. YANG, and C.-h. LI, *Deep Learning Based Visual Tracking: A Review*. DEStech Publications, Inc., 07 2017.

[39] P. Lynch, *The Emergence of Numerical Weather Prediction: Richardson's Dream*. Cambridge University Press, 2006.

[40] E. N. Parker, *Cosmical Magnetic Fields: Their Origin and their Activity (The International Series of Monographs on Physics)*. Oxford University Press, 1979.

[41] E. Sabine, "On Periodical Laws Discoverable in the Mean Effects of the Larger Magnetic Disturbances," *Philosophical Transactions of the Royal Society of London Series I*, vol. 141, pp. 123–139, 1851.

[42] E. Sabine, "On Periodical Laws Discoverable in the Mean Effects of the Larger Magnetic Disturbances. No. II," *Philosophical Transactions of the Royal Society of London Series I*, vol. 142, pp. 103–124, 1852.

[43] S. T. Suess, "The solar wind during the Maunder minimum," *Planet. Space Sci.*, vol. 27, pp. 1001–1013, Jul. 1979.

[44] J. A. Eddy, "The Maunder Minimum - A reappraisal," *Sol. Phys.*, vol. 89, pp. 195–207, Nov. 1983.

[45] E. N. Parker, "The passage of energetic charged particles through interplanetary space," *Planet. Space Sci.*, vol. 13, pp. 9–49, Jan. 1965.

[46] D. C. Wilkinson, M. A. Shea, and D. F. Smart, "A Case History of Solar and Galactic Space Weather Effects on the Geosynchronous Communications Satellite TDRS-1," *Advances in Space Research*, vol. 26, pp. 27–30, 2000.

[47] E. S. Babayev, "Some results of investigations on the space weather influence on functioning of several engineering-technical and communication systems and human health," *Astronomical and Astrophysical Transactions*, vol. 22, pp. 861–867, Jun. 2003.

[48] K. Schatten, "Fair space weather for solar cycle 24," *Geophys. Res. Lett.*, vol. 32, p. L21106, Nov. 2005.

[49] J. G. Kappenman, "An overview of the impulsive geomagnetic field disturbances and power grid impacts associated with the violent Sun-Earth connection events of 29-31 October 2003 and a comparative evaluation with other contemporary storms," *Space Weather*, vol. 3, p. S08C01, Aug. 2005.

[50] R. E. Turner, "Space Weather Challenges Intrinsic to the Human Exploration of Space," *Washington DC American Geophysical Union Geophysical Monograph Series*, vol. 165, p. 367, 2006.

[51] D. H. Hathaway and R. M. Wilson, "Geomagnetic activity indicates large amplitude for sunspot cycle 24," *Geophys. Res. Lett.*, vol. 33, p. L18101, Sep. 2006.

[52] G. Cornélissen, R. Tarquini, F. Perfetto, K. Otsuka, M. Gigolashvili, and F. Halberg, "Investigation of Solar about 5-Month Cycle in Human Circulating Melatonin: Signature of Weather in Extraterrestrial Space?" *Sun and Geosphere*, vol. 4, pp. 55–59, Dec. 2009.

[53] H.-S. Choi, J. Lee, K.-S. Cho, Y.-S. Kwak, I.-H. Cho, Y.-D. Park, Y.-H. Kim, D. N. Baker, G. D. Reeves, and D.-K. Lee, "Analysis of GEO spacecraft anomalies: Space weather relationships," *Space Weather*, vol. 9, p. 06001, Jun. 2011.

[54] M. West, D. Seaton, M. Dominique, D. Berghmans, B. Nicula, E. Pylyser, K. Stegen, and J. De Keyser, "Space Weather and Particle Effects on the Orbital Environment of PROBA2," in *EGU General Assembly Conference Abstracts*, ser. EGU General Assembly Conference Abstracts, vol. 15, Apr. 2013, pp. EGU2013–10 865.

[55] C. J. Schrijver, "Socio-Economic Hazards and Impacts of Space Weather: The Important Range Between Mild and Extreme," *Space Weather*, vol. 13, pp. 524–528, Sep. 2015.

[56] P. Grassberger, "Generalizations of the Hausdorff dimension of fractal measures," *Physics Letters A*, vol. 107, pp. 101–105, Jan. 1985.

[57] R. Bellman, *Dynamic Programming (Dover Books on Computer Science)*. Dover Publications, 2003.

[58] D. Cox and N. Pinto, "Beyond simple features: A large-scale feature search approach to unconstrained face recognition," in *Face and Gesture 2011*, March 2011, pp. 8–15.

[59] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," in *Readings in Speech Recognition*, A. Waibel and K.-F. Lee, Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990, pp. 393–404.

[60] K. Luk, J. E. Ball, and A. Sharma, "A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting," *Journal of Hydrology*, vol. 227, no. 1, pp. 56–65, 2000.

[61] R. J. Frank, N. Davey, and S. P. Hunt, "Input window size and neural network predictors," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, vol. 2, 2000, pp. 237–242 vol.2.

[62] R. J. Frank, N. Davey, and S. P. Hunt, "Time series prediction and neural networks," *Journal of Intelligent and Robotic Systems*, vol. 31, no. 1, pp. 91–103, 2001.

[63] K. J. Oh and K. jae Kim, "Analyzing stock market tick data using piecewise nonlinear model," *Expert Systems with Applications*, vol. 22, no. 3, pp. 249 – 255, 2002.

[64] Z. Sheng, L. Hong-Xing, G. Dun-Tang, and D. Si-Dan, "Determining the input dimension of a neural network for nonlinear time series prediction," *Chinese Physics*, vol. 12, no. 6, p. 594, 2003.

[65] A. H. Ghaderi, B. Bharani, and H. Jalalkamali, "Embedding dimension as input dimension of artificial neural network: A study on stock prices time series." *International Journal of Modern Physics and Applications*, vol. 1, no. 3, pp. 64–72, 2015.

[66] F. Takens, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics, Berlin Springer Verlag*, vol. 898, p. 366, 1981.

[67] H. Whitney, "Differentiable manifolds," *Ann. Math. (2)*, vol. 37, no. 3, pp. 645–680, July 1936, mR:1503303. Zbl:0015.32001. JFM:62.1454.01.

[68] R. Mañé, "On the dimension of the compact invariant sets of certain non-linear maps," *Lecture Notes in Mathematics, Berlin Springer Verlag*, vol. 898, p. 230, 1981.

[69] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *Journal of Statistical Physics*, vol. 65, pp. 579–616, Nov. 1991.

[70] H. Kantz and T. Schreiber, *Nonlinear time series analysis*, ser. Cambridge nonlinear science series. Cambridge, New York: Cambridge University Press, 1997, originally published: 1997.

[71] H. Abarbanel, *Analysis of Observed Chaotic Data*, ser. Institute for Nonlinear Science. Springer New York, 1997.

[72] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, pp. 1134–1140, Feb. 1986.

[73] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, pp. 3403–3411, Mar. 1992.

[74] J. M. Martinerie, A. M. Albano, A. I. Mees, and P. E. Rapp, "Mutual information, strange attractors, and the optimal estimation of dimension," *Phys. Rev. A*, vol. 45, pp. 7058–7064, May 1992.

[75] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, "The analysis of observed chaotic data in physical systems," *Reviews of Modern Physics*, vol. 65, pp. 1331–1392, Oct. 1993.

[76] H. D. I. Abarbanel and J. P. Gollub, "Analysis of Observed Chaotic Data," *Physics Today*, vol. 49, p. 86, Nov. 1996.

[77] M. Annunziato, S. Pizzuti, and L. S. Tsimring, "Analysis and prediction of spatio-temporal flame dynamics," in *Proceedings of the IEEE Workshop on Nonlinear Dynamics of Electronic Systems: Catania, Italy, 18-20 May 2000*, G. Setti, R. Rovatti, and G. Mazzini, Eds. World Scientific, 2000, pp. 117–121.

[78] A. Gkana and L. Zachilas, "Sunspot numbers: Data analysis, predictions and economic impacts," *Journal of Engineering Science and Technology Review*, vol. 8, no. 1, pp. 79–85, 2015, cited By 1.

[79] L. Zachilas and A. Gkana, "On the verge of a grand solar minimum: A second maunder minimum?" *Solar Physics*, vol. 290, no. 5, pp. 1457–1477, 2015.

[80] Y. Sun, V. Babovic, and E. S. Chan, "Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory," *Journal of Hydrology*, vol. 395, no. 12, pp. 109 – 116, 2010.

[81] S.-C. Huang, P.-J. Chuang, C.-F. Wu, and H.-J. Lai, "Chaos-based support vector regressions for exchange rate forecasting," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8590 – 8598, 2010.

[82] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537–550, Jul 1994.

[83] D. R. Kulkarni, A. S. Pandya, and J. C. Parikh, "Modeling and predicting sunspot activity-state space reconstruction + artificial neural network methods," *Geophys. Res. Lett.*, vol. 25, pp. 457–460, 1998.

[84] S. BuHamra, N. Smaoui, and M. Gabr, "The boxjenkins analysis and neural networks: prediction and time series modelling," *Applied Mathematical Modelling*, vol. 27, no. 10, pp. 805 – 815, 2003.

[85] R. Chandra and M. Zhang, "Cooperative coevolution of elman recurrent neural networks for chaotic time series prediction," *Neurocomputing*, vol. 86, pp. 116–123, 2012.

[86] Y. S. Maslennikova and V. V. Bochkarev, "Training algorithm for neuro-fuzzy network based on singular spectrum analysis," *CoRR*, vol. abs/1410.1151, 2014.

[87] T. Sauter, B. Weitzenkamp, and C. Schneider, "Spatio-temporal prediction of snow cover in the black forest mountain range using remote sensing and a recurrent neural network," *International Journal of Climatology*, vol. 30, no. 15, pp. 2330–2341, 2010.

[88] C. Jiang and F. Song, "Sunspot forecasting by using chaotic time-series analysis and narx network." *JCP*, vol. 6, no. 7, pp. 1424–1429, 2011.

[89] D. R. Kulkarni, J. C. Parikh, and A. S. Pandya, "Dynamic Predictions from Time Series Data - An Artificial Neural Network Approach," *International Journal of Modern Physics C*, vol. 8, pp. 1345–1360, 1997.

[90] P. Verdes, M. Parodi, P. Granitto, H. Navone, R. Piacentini, and H. Ceccatto, "Predictions of the maximum amplitude for solar cycle 23 and its subsequent behavior using nonlinear methods," *Solar Physics*, vol. 191, no. 2, pp. 419–425, 2000.

[91] F. Fessant, C. Pierret, and P. Lantos, "Comparison of Neural Network and McNish and Lincoln Methods for the Prediction of the Smoothed Sunspot Index," *Sol. Phys.*, vol. 168, pp. 423–433, Oct. 1996.

[92] P. S. Lucio, F. C. Conde, I. F. A. Cavalcanti, A. I. Serrano, A. M. Ramos, and A. O. Cardoso, "Spatiotemporal monthly rainfall reconstruction via artificial neural network - case study: south of Brazil," *Advances in Geosciences*, vol. 10, pp. 67–76, Apr. 2007.

[93] R. Chandra and M. Zhang, "Cooperative coevolution of elman recurrent neural networks for chaotic time series prediction," *Neurocomput.*, vol. 86, pp. 116–123, Jun. 2012.

[94] R. Archana, A. Unnikrishnan, and R. Gopikakumari, "Computation of state space evolution of chaotic systems from time series of output, based on neural network models," *International Journal of Engineering Research and Development*, vol. 2, pp. 49–56, Jul. 2012.

[95] P. Maass, T. Koehler, J. Kalden, R. Costa, U. Parlitz, C. Merkwirth, and J. Wichard, *Mathematical Methods for Forecasting Bank Transaction Data*, ser. DFG-Schwerpunktprogramm 1114, Mathematical methods for time series analysis and digital image processing. Zentrum für Technomathematik, 2003.

[96] H. D. Navone and H. A. Ceccatto, "Forecasting chaos from small data sets: a comparison of different nonlinear algorithms," *Journal of Physics A: Mathematical and General*, vol. 28, no. 12, p. 3381, 1995.

[97] R. A. Calvo, H. A. Ceccato, and R. D. Piacentini, "Neural network prediction of solar activity," *ApJ*, vol. 444, pp. 916–921, May 1995.

[98] G. Simon and M. Verleysen, "High-dimensional delay selection for regression models with mutual information and distance-to-diagonal criteria," *Neurocomput.*, vol. 70, no. 7-9, pp. 1265–1275, Mar. 2007.

[99] M. Ragulskis and K. Lukoseviciute, "Non-uniform attractor embedding for time series forecasting by fuzzy inference systems," *Neurocomputing*, vol. 72, no. 10, pp. 2618 – 2626, 2009, lattice Computing and Natural Computing (JCIS 2007) / Neural Networks in Intelligent Systems Designn (ISDA 2007).

[100] F. Liu, C. Quek, and G. S. Ng, "Neural network model for time series prediction by reinforcement learning," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, vol. 2, July 2005, pp. 809–814 vol. 2.

[101] P. J. Werbos, "Applications of advances in nonlinear sensitivity analysis," in *System Modeling and Optimization*, R. F. Drenick and F. Kozin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1982, pp. 762–770.

[102] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.

[103] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct 1990.

[104] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 1998, pp. 2278–2324.

[105] R. Reed and R. J. Marks II, *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks (MIT Press)*. A Bradford Book, 1999.

[106] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 13, no. 4, pp. 600–612, 2004.

[107] C. Dong, C. Change Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *ArXiv e-prints*, Dec. 2015.

[108] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing Data Reconstruction in Remote Sensing image with a Unified Spatial-Temporal-Spectral Deep Convolutional Neural Network," *ArXiv e-prints*, Feb. 2018.

[109] Z. Wang and A. C. Bovik, "Mean squared error: Lot it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Processing Magazine*, vol. 26, pp. 98–117, Jan. 2009.

[110] D. Brunet, E. R. Vrscay, and Z. Wang, "On the Mathematical Properties of the Structural Similarity Index," *IEEE Transactions on Image Processing*, vol. 21, pp. 1488–1499, Apr. 2012.

[111] E. F. Krause, *Taxicab Geometry: An Adventure in Non-Euclidean Geometry (Dover Books on Mathematics)*. Dover Publications, 1987.

[112] E. Covas, R. Tavakol, P. Ashwin, A. Tworkowski, and J. M. Brooke, "In-out intermittency in partial differential equation and ordinary differential equation models," *Chaos*, vol. 11, pp. 404–409, Jun. 2001.

[113] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*. Dover books on mathematics (Mineola, NY: Dover Publications), ISBN 0486411834, 2001.

[114] E. W. Maunder, "Note on the distribution of sun-spots in heliographic latitude, 1874-1902," *MNRAS*, vol. 64, pp. 747–761, Jun. 1904.

[115] "bfly.jpg (jpeg image, 8192 x 4358 pixels)," Mar 2018, [Online; accessed 20. Apr. 2018].

[116] B. Owens, "Long-term research: Slow science," *Nature*, vol. 495, pp. 300–303, Mar. 2013.

[117] H. C. Koons and D. J. Gorney, "A sunspot maximum prediction using a neural network," *EOS Transactions*, vol. 71, p. 677, May 1990.

[118] A. S. Weigend, "Connectionist Architectures for Time Series Prediction of Dynamical Systems," Ph.D. dissertation, STANFORD UNIVERSITY., 1991.

[119] A. S. Weigend, B. A. Huberman, and D. E. Rumelhart, "Predicting Sunspots and Exchange Rates with Connectionist Networks," in *Nonlinear modeling and forecasting*, M. Casdagli and S. Eubank, Eds. Addison-Wesley, 1992, pp. 395–432.

[120] K. MacPherson, "Neural network computation techniques applied to solar activity prediction," *Advances in Space Research*, vol. 13, pp. 447–450, Sep. 1993.

[121] K. P. Macpherson, A. J. Conway, and J. C. Brown, "Prediction of solar and geomagnetic activity data using neural networks," *J. Geophys. Res.*, vol. 100, pp. 21 735–21 744, Nov. 1995.

[122] A. Conway, "Echoed time series predictions, neural networks and genetic algorithms," *Vistas in Astronomy*, vol. 38, pp. 351–356, 1994.

[123] T. Koskela, M. Lehtokangas, J. Saarinen, and K. Kaski, "Time series prediction with multilayer perceptron, fir and elman neural networks," in *In Proceedings of the World Congress on Neural Networks*. Press, 1996, pp. 491–496.

[124] F. Fessant, S. Bengio, and D. Collobert, "On the prediction of solar activity using different neural network models," *Annales Geophysicae*, vol. 14, pp. 20–26, Jan. 1996.

[125] Y. R. Park, T. J. Murray, and C. Chen, "Predicting sunspots using a layered perceptron neural network." *IEEE Transactions on Neural Networks*, vol. 7, pp. 501–505, Mar. 1996.

[126] A. J. Conway, K. P. Macpherson, G. Blacklaw, and J. C. Brown, "A neural network prediction of solar cycle 23," *J. Geophys. Res.*, vol. 103, pp. 29 733–29 742, Dec. 1998.

[127] A. J. Conway, "Time series, neural networks and the future of the Sun," *New A Rev.*, vol. 42, pp. 343–394, Oct. 1998.

[128] P. F. Verdes, P. M. Granitto, and H. A. Ceccatto, "Secular Behavior of Solar Magnetic Activity: Nonstationary Time-Series Analysis of the Sunspot Record," *Sol. Phys.*, vol. 221, pp. 167–177, May 2004.

[129] H. Lundstedt, "Solar activity predicted with artificial intelligence," *Washington DC American Geophysical Union Geophysical Monograph Series*, vol. 125, pp. 201–204, 2001.

[130] M. Small and C. K. Tse, "Minimum description length neural networks for time series prediction," *Phys. Rev. E*, vol. 66, no. 6, p. 066701, Dec. 2002.

[131] A. V. Mordvinov, N. G. Makarenko, M. G. Ogurtsov, and H. Jungner, "Reconstruction of Magnetic Activity of the Sun and Changes in Its Irradiance on a Millennium Timescale Using Neurocomputing," *Sol. Phys.*, vol. 224, pp. 247–253, Oct. 2004.

[132] A. Gholipour, C. Lucas, B. N. Araabi, and M. Shafiee, "Solar activity forecast: Spectral analysis and neurofuzzy prediction," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 67, pp. 595–603, Apr. 2005.

[133] A.-F. A. Attia, R. H. Abdel-Hamid, and M. Quassim, "A genetic-based neuro-fuzzy approach for prediction of solar activity," in *Modeling and Systems Engineering for Astronomy*, ser. Proc. SPIE, S. C. Craig and M. J. Cullum, Eds., vol. 5497, Sep. 2004, pp. 542–552.

[134] M. S. Quassim and A. F. Attia, "Forecasting the global temperature trend according to the predicted solar activity during the next decades," *Mem. Soc. Astron. Italiana*, vol. 76, p. 1030, 2005.

[135] A.-F. Attia, R. Abdel-Hamid, and M. Quassim, "Prediction of Solar Activity Based on Neuro-Fuzzy Modeling," *Sol. Phys.*, vol. 227, pp. 177–191, Mar. 2005.

[136] M. Mirmomeni, M. Shafiee, C. Lucas, and B. N. Araabi, "Introducing a new learning method for fuzzy descriptor systems with the aid of spectral analysis to forecast solar activity," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 68, pp. 2061–2074, Dec. 2006.

[137] M. S. Quassim, A.-F. Attia, and H. K. Elminir, "Forecasting the Peak Amplitude of the 24th and 25th Sunspot Cycles and Accompanying Geomagnetic Activity," *Sol. Phys.*, vol. 243, pp. 253–258, Jul. 2007.

[138] A.-F. Attia, H. A. Ismail, and H. M. Basurah, "A Neuro-Fuzzy modeling for prediction of solar cycles 24 and 25," *Ap&SS*, vol. 344, pp. 5–11, Mar. 2013.

[139] J.-X. Xie, C.-T. Cheng, K.-W. Chau, and Y.-Z. Pei, "A hybrid adaptive time-delay neural network model for multi-step-ahead prediction of sunspot activity," *International Journal of Environment and Pollution*, vol. 28, no. 3-4, pp. 364–381, 2006.

[140] G. Maris and A. Oncica, "Solar Cycle 24 Forecasts," *Sun and Geosphere*, vol. 1, no. 1, pp. 8–11, Mar. 2006.

[141] F. Emmert-Streib and M. Dehmer, "Nonlinear Time Series Prediction Based on a Power-Law Noise Model," *International Journal of Modern Physics C*, vol. 18, pp. 1839–1852, 2007.

[142] A. S. Pandya, D. R. Kulkarni, and J. C. Parikh, "Study of time series prediction under noisy environment," in *Applications and Science of Artificial Neural Networks III*, ser. Proc. SPIE, S. K. Rogers, Ed., vol. 3077, Apr. 1997, pp. 116–126.

[143] H. Lundstedt, M. Wik, and P. Wintoft, "Synoptic Solar Magnetic Fields: Explored and Predicted," *AGU Fall Meeting Abstracts*, Dec. 2006.

[144] M. Wik, "Multiresolution Analysis and Prediction of Solar Magnetic Flux," in *37th COSPAR Scientific Assembly*, ser. COSPAR Meeting, vol. 37, 2008, p. 3467.

[145] D. Gang and Z. Shi-Sheng, "Sunspot number prediction based on process neural network with time-varying threshold functions [j]," *Acta Physica Sinica*, vol. 2, p. 099, 2007.

[146] J. Uwamahoro, L.-A. McKinnell, and P. J. Cilliers, "Forecasting solar cycle 24 using neural networks," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 71, pp. 569–574, Apr. 2009.

[147] C. Francile and M. L. Luoni, "Hacia la predicción del Número R de Wolf de manchas solares utilizando Redes Neuronales con retardos temporales," *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina*, vol. 53, pp. 241–244, 2010.

[148] M. A. Parodi, H. A. Ceccatto, R. D. Piacentini, and P. J. García, "Actividad solar del ciclo 23. Predicción del máximo y fase decreciente utilizando redes neuronales," *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina*, vol. 43, pp. 23–24, 1999.

[149] A. Ajabshirizadeh, N. Masoumzadeh Jouzdani, and S. Abbassi, "Neural network prediction of solar cycle 24," *Research in Astronomy and Astrophysics*, vol. 11, pp. 491–496, Apr. 2011.

[150] A. Ajabshirizadeh and M. Juzdani Nafiseh, "Sunspot Number and Solar Radio Flux Prediction by Artificial Neural Network Method." in *38th COSPAR Scientific Assembly*, ser. COSPAR Meeting, vol. 38, 2010, p. 2.

[151] S. Chattopadhyay, D. Jhajharia, and G. Chattopadhyay, "Trend estimation and univariate forecast of the sunspot numbers: Development and comparison of ARMA, ARIMA and Autoregressive Neural Network models," *Comptes Rendus Geoscience*, vol. 343, pp. 433–442, Jul. 2011.

[152] Y. Maslennikova and V. Bochkarev, "Solar activity prediction using artificial neural network and singular spectrum analysis," in *39th COSPAR Scientific Assembly*, ser. COSPAR Meeting, vol. 39, Jul. 2012, p. 1194.

[153] D.-C. Park and D.-M. Woo, "Prediction of sunspot series using bilinear recurrent neural network," in *Information Management and Engineering, 2009. ICIME'09. International Conference on*. IEEE, 2009, pp. 94–98.

[154] T.-H. Kim, D.-C. Park, D.-M. Woo, W. Huh, C.-H. Yoon, H.-U. Kim, and Y. Lee, "Sunspot series prediction using a multiscale recurrent neural network," in *Signal Processing and Information Technology (ISSPIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 399–403.

[155] J. D. Moghaddam, A. Mosallanezhad, and M. Teshnehlab, "Sunspot prediction by a time delay line recurrent fuzzy neural network using emotional learning," in *Fuzzy Systems (IFSC), 2013 13th Iranian Conference on*. IEEE, 2013, pp. 1–5.

[156] G. Chattopadhyay and S. Chattopadhyay, "Monthly sunspot number time series analysis and its modeling through autoregressive artificial neural network," *European Physical Journal Plus*, vol. 127, p. 43, Apr. 2012.

[157] Z.-G. Liu and J. Du, "Sunspot time sequences prediction based on process neural network and quantum particle swarm," in *Multimedia Information Networking and Security (MINES), 2012 Fourth International Conference on*. IEEE, 2012, pp. 233–236.

[158] M. Parsapoor, U. Bilstrup, and B. Svensson, "Prediction of solar cycle 24," in *2015 International Joint Conference on Neural Networks, IJCNN 2015, Killarney, Ireland, July 12-17, 2015*. IEEE, 2015, pp. 1–8.

[159] M. Parsapoor, J. Brooke, and B. Svensson, "A new computational intelligence model for long-term prediction of solar and geomagnetic activity," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA.*, B. Bonet and S. Koenig, Eds. AAAI Press, 2015, pp. 4192–4193.

[160] J. Jiang, R. H. Cameron, D. Schmitt, and M. Schüssler, "The solar magnetic field since 1700. I. Characteristics of sunspot group emergence and reconstruction of the butterfly diagram," *A&A*, vol. 528, p. A82, Apr. 2011.

[161] R. H. Cameron, J. Jiang, and M. Schüssler, "Solar Cycle 25: Another Moderate Cycle?" *ApJ*, vol. 823, p. L22, Jun. 2016.

[162] S. W. McIntosh, X. Wang, R. J. Leamon, A. R. Davey, R. Howe, L. D. Krista, A. V. Malanushenko, R. S. Markel, J. W. Cirtain, J. B. Gurman, W. D. Pesnell, and M. J. Thompson, "Deciphering solar magnetic activity. i. on the relationship between the sunspot cycle and the evolution of small magnetic features," *The Astrophysical Journal*, vol. 792, no. 1, p. 12, 2014.

[163] J. Jiang and J. Cao, "Predicting solar surface large-scale magnetic field of Cycle 24," *ArXiv e-prints*, Jul. 2017.

[164] N. Safiullin, N. Kleeorin, S. Porshnev, I. Rogachevskii, and A. Ruzmaikin, "Nonlinear mean-field dynamo and prediction of solar activity," *ArXiv e-prints*, Dec. 2017.

[165] K. Kaneko, "Towards Thermodynamics of Spatiotemporal Chaos," *Progress of Theoretical Physics Supplement*, vol. 99, pp. 263–287, 1989.
[166] G. Mayer-Kress and K. Kaneko, "Spatiotemporal chaos and noise," *Journal of Statistical Physics*, vol. 54, pp. 1489–1508, Mar. 1989.
[167] K. Kaneko, *Theory and Applications of Coupled Map Lattices (Nonlinear Science: Theory and Applications)*. Wiley, 1993.
[168] E. Lorenz, "Predictability: a problem partly solved," in *Seminar on Predictability, 4-8 September 1995*, vol. 1, ECMWF. Shinfield Park, Reading: ECMWF, 1995, pp. 1–18.
[169] "Lorenz '96 model - File Exchange - MATLAB Central," Apr 2018, [Online; accessed 20. Apr. 2018].
[170] Y. Kuramoto and T. Tsuzuki, "Persistent Propagation of Concentration Waves in Dissipative Media Far from Thermal Equilibrium," *Progress of Theoretical Physics*, vol. 55, pp. 356–369, Feb. 1976.
[171] G. I. Sivashinsky, "Nonlinear analysis of hydrodynamic instability in laminar flames I. Derivation of basic equations," *Acta Astronautica*, vol. 4, pp. 1177–1206, 1977.
[172] "Kuramoto-Sivashinsky: an investigation of spatiotemporal "turbulence"," Apr 2007, [Online; accessed 20. Apr. 2018].

Eurico Covas Biography text here.

Emmanouil Benetos Biography text here.