
Control-aware echo state networks (Ca-ESN) for the suppression of extreme events

Alberto Racca

I-X and Department of Aeronautics
Imperial College London
a.racca@imperial.ac.uk

Luca Magri

Department of Aeronautics
Imperial College London,
The Alan Turing Institute
l.magri@imperial.ac.uk

Abstract

Extreme events are sudden large-amplitude changes in the state or observables of chaotic nonlinear systems, which characterize many scientific phenomena. Because of their violent nature, extreme events typically have adverse consequences, which call for methods to prevent the events from happening. In this work, we introduce the control-aware echo state network (Ca-ESN) to seamlessly combine ESNs and control strategies, such as proportional-integral-derivative and model predictive control, to suppress extreme events. The methodology is showcased on a chaotic-turbulent flow, in which we reduce the occurrence of extreme events with respect to traditional methods by two orders of magnitude. This work opens up new possibilities for the efficient control of nonlinear systems with neural networks.

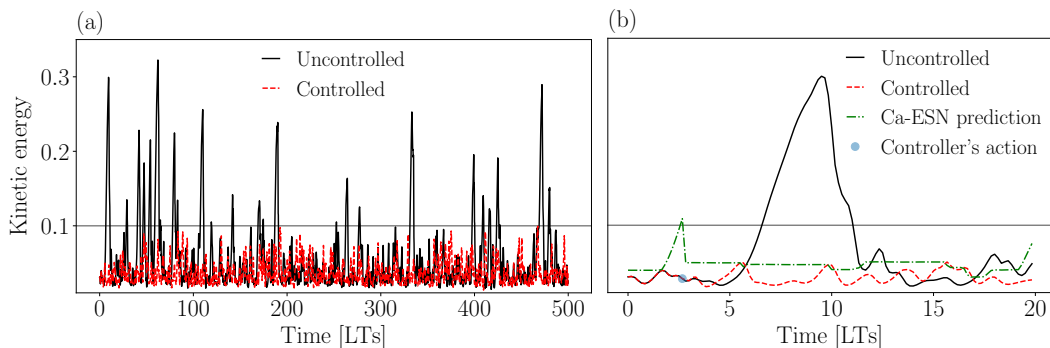


Figure 1: (a) Controlled and uncontrolled system's evolution. (b) Example of a suppressed extreme event. The controller acts before the event occurs based on the echo state network's prediction.

1 Introduction

Extreme events arise in a variety of natural and engineering systems, in the form of rogue waves, atmospheric events and power grid shocks, to name a few [6]. These events often have negative consequences, thus, developing methods for their prediction and control is an active field of research [6, 16]. At the same time, the events usually show no apparent early sign of their occurrence, which makes their control from direct observations especially difficult. Because of this, control of extreme events is typically performed by assuming knowledge of the governing equations of the system, whose prediction via time integration activates the controller in advance [12]. When the governing

equations are not known, data-driven modelling becomes necessary. To this end, echo state networks (ESNs) [8], which are state-of-the-art machines for the prediction of nonlinear dynamical systems, offer a promising alternative [19]. ESNs have been shown to predict extreme events, and deployed to activate basic controllers for their suppression [13, 15]. These works, however, considered only simplistic prescribed control strategies, which have limited the applicability and performance of the controllers. We propose the control-aware ESN (Ca-ESN) to seamlessly integrate the prediction of ESNs in the formalism of established controllers, such as proportional integral derivative (PID) controllers [1] and model predictive control (MPC) [4, 5], to efficiently suppress extreme events. We showcase the Ca-ESN in a model of turbulence [10], in which extreme events occur as intermittent burst in the total kinetic energy of the flow.

2 The control problem: suppression of extreme events

We analyse dynamical systems that evolve according to a set of discretized governing equations

$$\mathbf{q}(t_{i+1}) = \mathbf{f}(\mathbf{q}(t_i), \mathbf{u}(t_i)), \quad (1)$$

where $\mathbf{q}(t_i)$ is the state of the system, which shows extreme events in the observable, $k(t_i) = \mathbf{g}(\mathbf{q}(t_i))$ (Fig. 1). We assume that \mathbf{f} is not known, and that data on \mathbf{q} is available. We wish to suppress the events through the control input, $\mathbf{u}(t_i)$, whose objective is to decrease the number of extreme events while acting as infrequently as possible on the system. This is translated into a quantitative goal through the average reward

$$R = \frac{1}{N} \sum_{i=1}^N r(t_i), \quad (2)$$

where $r(t_i)$ is the user-defined reward at each time step. First, to prevent extreme events from happening, we set a negative reward, $r(t_{i,e}) = -1$, for each time step, $t_{i,e}$, at which the system is experiencing an event. Second, to find control strategies that act rarely on the system, we select a system-dependent (smaller) negative reward, $r(t_{i,c})$, for every time step, $t_{i,c}$, in which the control strategy is activated. This is to discourage the activation of the control strategy when it is not needed. The reward is set to zero for all other time steps. To further characterize the controlled system, we compute the total number of time steps that the system experiences an extreme event, N_e , or control, N_c . By doing so, we analyse how often (i) the system shows extreme behaviour through the extreme events ratio, P_e , and (ii) control is active through the control ratio, P_c ,

$$P_e = \frac{N_e}{N}, \quad P_c = \frac{N_c}{N}. \quad (3)$$

3 Control-aware echo state network

To suppress extreme events, we employ standard controllers. We first consider the proportional-integral-derivative (PID) controller [1], $c(k(t))$,

$$c(k(t)) = K_p k(t) + K_d \frac{dk(t)}{dt} + K_i \int_{t-\tau_i}^t k(t') dt'; \quad (4)$$

where the proportional, K_p , derivative, K_d and integral, K_i , multipliers and the integral time, τ_i , are selected through Bayesian optimisation [17] to optimize the average reward. Secondly, we analyse nonlinear model predictive control (MPC), which finds the optimal control sequence within a future time window [4]. In MPC, at every control step, t_0 , we solve a constrained optimization problem to maximize the average reward over the future (receding) time horizon, $\tau_{\text{hor}} = N_{\text{hor}} dt$,

$$\max_{\mathbf{u}_{\text{opt}}(t_i)} R_{\text{hor}} = \frac{1}{N_{\text{hor}}} \sum_{i=1}^{N_{\text{hor}}} r(t_i)$$

subject to $\mathbf{q}(t_{i+1}) = \mathbf{f}(\mathbf{q}(t_i), \mathbf{u}_{\text{opt}}(t_i))$ for $i \leq N_{\text{opt}}$

$$\mathbf{q}(t_{i+1}) = \mathbf{f}(\mathbf{q}(t_i), \mathbf{u}_{\text{fix}}(t_i)), \quad \text{for } N_{\text{opt}} < i < N_{\text{hor}} \quad (5)$$

where $N_{\text{opt}} \leq N_{\text{hor}}$ are the time steps at which we optimise the control law, $[\mathbf{u}_{\text{opt}}(t_0), \dots, \mathbf{u}_{\text{opt}}(t_0 + N_{\text{opt}} dt)]$ [11]. The system is controlled for the remaining time steps within the time horizon by a

prescribed (fixed) control law, $\mathbf{u}_{\text{fix}} = \mathbf{0}$ (no control). We use two different strategies, \mathbf{u}_{fix} and \mathbf{u}_{opt} , to decrease the search space and therefore the computational cost of solving (5).

To enable the controllers, we propose the Control-aware Echo State Network (Ca-ESN). Echo state networks [8] nonlinearly expand the inputs into a high-dimensional reservoir, $\mathbf{r}(t_i)$, from which the output of the network is computed as a linear combination. The Ca-ESN provides a data-driven model for the iterative evolution of the controlled system, $\hat{\mathbf{q}}(t_{i+1}) = \mathbf{f}_{\text{ESN}}(\mathbf{q}(t_i), \mathbf{u}(t_i))$,

$$\begin{aligned} \mathbf{r}(t_{i+1}) &= \tanh(\sigma_{\text{in}} \mathbf{W}_{\text{in}} \mathbf{q}_{\text{in}}(t_i) + \rho \mathbf{W} \mathbf{r}(t_i) + \sigma_c \mathbf{W}_c \mathbf{u}(t_i)), \\ \hat{\mathbf{q}}(t_{i+1}) &= \mathbf{r}(t_{i+1})^T \mathbf{W}_{\text{out}}; \end{aligned} \quad (6)$$

where the matrices \mathbf{W}_{in} , \mathbf{W} , and \mathbf{W}_c are randomly generated and fixed [9]. The hyperparameters σ_{in} and σ_c are optimized through Bayesian optimisation and recycle validation [14]. Because the evolution of the system is Markovian (1), we set $\rho = 0$ to eliminate the recurrence in time of the network. In this way, we simplify the architecture, thereby (i) reducing its computational cost and (ii) making it equivalent to a one-layer extreme learning machine [7]. The weights of the output matrix, \mathbf{W}_{out} , are the only trainable parameters. Thanks to this architecture, training the network needs only solving a ridge regression problem [9]. In this way, training does not require either backpropagation or gradient descent, which is usually problematic in time series forecasting [2].

Once the networks are trained, we integrate the Ca-ESN predictions in the formalism of controllers (4)-(5). In the PID controller, we use the maximum of the predicted observable for the uncontrolled system in the horizon τ_{hor} , as the control variable, $k(t) = \max(\hat{k}_{\text{ESN}}(t, t + \tau_{\text{hor}}))$. Because we use the maximum, we simplify the controller by neglecting the derivative and integral terms ($K_d = K_i = 0$). A schematic implementation is shown in Algorithm 1. In MPC, the network provides the model: $\mathbf{f}_{\text{ESN}}(\mathbf{q}(t_i), \mathbf{u}(t_i))$ predicts the future evolution of the controlled system in (5)¹.

Finally, we remark that the current methodology can be extended to other control strategies, such as reinforcement learning [18]. This can be achieved by employing the Ca-ESN prediction as part of the observations on which the control strategy bases its decision.

Algorithm 1 PID control with the Ca-ESN

```

for  $t \leftarrow t_0$  to  $t_N$  do                                     ▷ Every control step
     $\hat{\mathbf{q}}(t) = \mathbf{q}(t)$ 
     $k(t) = \mathbf{g}(\hat{\mathbf{q}}(t))$ 
    for  $t' \leftarrow t$  to  $t + \tau_{\text{hor}}$  do                             ▷ Evaluate the Ca-ESN up to  $\tau_{\text{hor}}$ 
         $\hat{\mathbf{q}}(t' + dt) = \mathbf{f}_{\text{ESN}}(\hat{\mathbf{q}}(t'), \mathbf{u}(t'))$ 
         $k(t) = \max(k(t), \mathbf{g}(\hat{\mathbf{q}}(t' + dt)))$                  ▷ Save maximum of the observable
    end for
     $\mathbf{u}(t) = c(k(t))$                                            ▷ Select control with PID
     $\mathbf{q}(t + dt) = \mathbf{f}(\mathbf{q}(t), \mathbf{u}(t))$                              ▷ Apply control
end for

```

4 Results

To test the Ca-ESN, we consider the MFE, which is a qualitative model of turbulence [10]. The dynamics are governed by the non-dimensional incompressible Navier-Stokes equations

$$\nabla \cdot \mathbf{v} = 0, \quad \frac{d\mathbf{v}}{dt} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla p + \frac{1}{\text{Re}} \Delta \mathbf{v} + \mathbf{f}, \quad (7)$$

where \mathbf{v} is the velocity, p is the pressure, Re is the Reynolds number, and \mathbf{f} is the body forcing that sustains turbulence. The MFE model is generated by projecting (7) onto compositions of Fourier modes, which spawns nine nonlinear ordinary differential equations for the amplitudes of the modes, $q_i(t)$, which become the unknowns of the system [10]. To integrate the equations, we use the same parameters and boundary conditions as [15]. For a wide range of Reynolds numbers, the system displays chaotic dynamics characterized by extreme events of the kinetic energy, $k(t) = \frac{1}{2} \sum_{i=1}^9 q_i^2(t)$; $k(t) > k_e = 0.1$ (Fig. 1, 2a), whose probability decreases with Re [15].

¹The model is implemented in JAX [3], and publicly available on GitHub.

We analyse the highly extreme $Re = 400$ regime, and suppress extreme events by temporarily increasing the Reynolds number, i.e. $\mathbf{u}(t_i) = Re(t_i) = 2000$, following [15]. We set the control penalisation term, $r(t_{i,c}) = -0.15$, for the two regimes to be equally desirable on average ($R_{400} \simeq R_{2000}$), so that a combination of the two is found by the controller. In this control setting, (i) the PID controller is activated when $c(t) > k_c$, where k_c is optimised instead of the proportional multiplier ($K_p = 1$), and (ii) the optimisation of the discrete MPC problem is solved through complete search.

We test the control strategies on 100,000 time series of length 20 Lyapunov times (LT)². One representative time series, in which the event is suppressed through the Ca-ESN proportional controller, is shown in Fig. 1b. The network is trained on 50 time series only, and evolves for an horizon $\tau_{\text{hor}} = 4$ LTs, with a control horizon $N_{\text{opt}}dt = 1$ LT, which are selected as a trade-off between the computational cost of the prediction and its capability of suppressing extreme events. The controllers act on the system every 10 time units, which result in $\mathcal{O}(10^7)$ control steps analysed for each strategy.

Figure 2 shows the quantitative results. First, the standard PID controller decreases the number of extreme events by more than one order of magnitude compared to the uncontrolled (NC) system (Fig. 2c). Second, integrating the Ca-ESN in the proportional controller (P_{ESN}) and model predictive control (MPC), markedly improves the reward of the controllers with respect to the literature (Lit) [15] (Fig. 2b). The Ca-ESN controllers decrease the occurrence of extreme events, while requiring significantly fewer actions than other methods (Fig. 2c,d). This shows that employing the networks is highly beneficial for the suppression of the events. Third, the Ca-ESN controllers decrease the occurrence of extreme events with respect to always controlling (AC) the system (Fig. 2c). This indicates that optimally selecting the active control strategy is more effective than passive control.

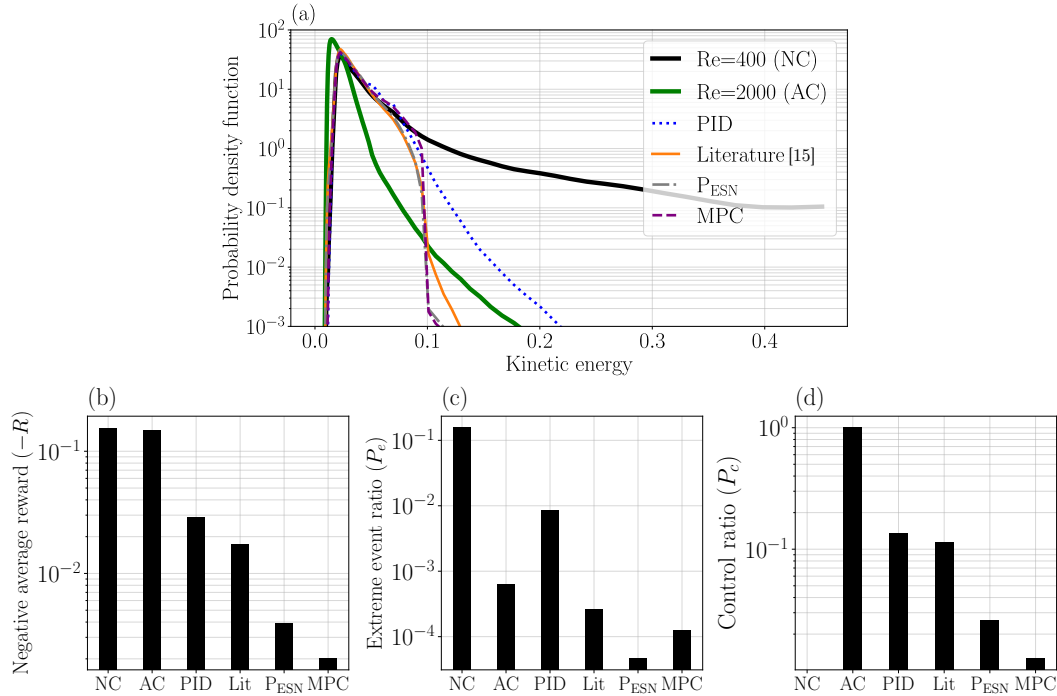


Figure 2: (a) Probability density function of the kinetic energy, (b) average reward, (c) extreme event ratio and (d) control ratio for different control strategies.

5 Conclusions

We propose the control-aware echo state network (Ca-ESN) to integrate ESNs into the formalism of conventional control strategies to suppress extreme events in chaos. The architecture is demonstrated

²The Lyapunov time is the inverse of the Lyapunov exponent of the system, which measures the average divergence of close-by trajectories in chaotic dynamics. In the MFE, $1\text{LT} = 0.0163^{-1}$ time units [15].

on a turbulent flow, in which we combine the networks with PID controllers and model predictive control. We show that the Ca-ESN (i) decreases the occurrence of extreme events up to two orders of magnitude with respect to both the uncontrolled and standard PID scenarios, (ii) requires an order of magnitude fewer actions to do so, and (iii) is more effective in suppressing the events than controlling the system at all times (passive control). This work opens up opportunities for the efficient control of extreme nonlinear dynamics from data, without the knowledge of the governing equations.

Acknowledgments and Disclosure of Funding

A. R. is supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program. L. M. gratefully acknowledges financial support from the ERC Starting Grant PhyCo 949388 and from the UKRI AI for Net Zero grant EP/Y005619/1.

References

- [1] Karl J Astrom. Pid controllers: theory, design, and tuning. *The International Society of Measurement and Control*, 1995.
- [2] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.
- [3] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. *JAX: composable transformations of Python+NumPy programs*, 2018.
- [4] Eduardo F Camacho and Carlos Bordons Alba. *Model predictive control*. Springer science & business media, 2013.
- [5] Andreas Draeger, Sebastian Engell, and Horst Ranke. Model predictive control using neural networks. *IEEE Control Systems Magazine*, 15(5):61–66, 1995.
- [6] Mohammad Farazmand and Themistoklis P. Sapsis. Extreme Events: Mechanisms and Prediction. *Applied Mechanics Reviews*, 71(5), 08 2019. 050801.
- [7] Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2):513–529, 2011.
- [8] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, 2004.
- [9] Mantas Lukoševičius. A practical guide to applying echo state networks. In *Neural networks: Tricks of the trade*, pages 659–686. Springer, 2012.
- [10] Jeff Moehlis, Holger Faisst, and Bruno Eckhardt. A low-dimensional model for turbulent shear flows. *New Journal of Physics*, 6(1):56, 2004.
- [11] Richard M Murray et al. Optimization-based control. *California Institute of Technology, CA*, pages 111–128, 2009.
- [12] Viktor Nagy and Edward Ott. Control of rare intense events in spatiotemporally chaotic systems. *Physical Review E*, 76(6):066206, 2007.
- [13] Viktoras Pyragas and Kestutis Pyragas. Using reservoir computer to predict and prevent extreme events. *Physics Letters A*, 384(24):126591, 2020.
- [14] Alberto Racca and Luca Magri. Robust optimization and validation of echo state networks for learning chaotic dynamics. *Neural Networks*, 142:252–268, 2021.
- [15] Alberto Racca and Luca Magri. Data-driven prediction and control of extreme events in a chaotic flow. *Phys. Rev. Fluids*, 7:104402, Oct 2022.
- [16] Themistoklis P Sapsis. Statistics of extreme events in fluid flows and waves. *Annual Review of Fluid Mechanics*, 53:85–111, 2021.
- [17] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

- [18] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [19] P.R. Vlachas, J. Pathak, B.R. Hunt, T.P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos. Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. *Neural Networks*, 126:191–217, 2020.