Predicting Extreme Wind Events in Colorado's Front Range: Reservoir Computing Approach

Wisang Sugiarta
Dept. of Computer Science
University of Colorado - Boulder

Andrew Winters
Dept. of Atmospheric and Oceanic Science
University of Colorado - Boulder

1 Introduction

The study of dynamical systems has been on the forefront of science since the late 1800s. It was first introduced by Henry Poincaré in his studies of celestial mechanics [1]. The systems can range from simple pendulums to high-dimensional models of the atmosphere. Poincaré and the field set out to study how the state of systems evolves over time, expressed as differential equations (DE). These differential equations are fundamental to understanding how a system evolves, and simulations can help understand the future.

Quickly, scientists found systems that exhibited obvious patterns but diverged significantly when initial changes were made. This came to be a staple of chaotic systems and known as the "butterfly effect" introduced by meteorologist Edward Lorenz in his groundbreaking work [2]. The current view of weather as chaotic comes from Lorenz's same paper. Many physicists have been developing the explicit DEs, but with the explosion of data availability, many algorithms to develop data-driven models have been created.

Extreme wind events in Boulder, Colorado, are a compelling example of such behavior, resulting from complex, nonlinear couplings between synoptic-scale flows, terrain-induced turbulence, and local weather dynamics [3]. This project explores whether different modeling approaches — both classical chaotic time-series analysis and modern reservoir computing — can effectively forecast these events and uncover structure in the underlying spatiotemporal chaos.

Reservoir computers are particularly well-suited to this task[4, 5, 6]: they offer a computationally efficient method of modeling temporal dynamics in high-dimensional, nonlinear systems while maintaining the flexibility of a data-driven approach. Given the availability of dense local weather observations from the NCAR Boulder Mesa station, and motivated by related atmospheric research into downslope windstorms, this study applies reservoir computing to short-term extreme wind prediction in Boulder. The goal is to assess whether such an approach can extract predictive signals from local atmospheric time series and operate effectively under data constraints.

2 Methodology

2.1 Data Description

The goal of the study is to optimize an algorithm to best predict extreme winds. We chose to pursue a data-driven approach, and an important component of the methodology is selecting a reliable, robust dataset. We decided to select the National Center for Atmospheric Research (NCAR)[7] Boulder weather site which may be found at

`https://archive.eol.ucar.edu/cgi-bin/weather.cgi?site=ml&units=english`

It is a reliable hourly dataset that dates back to 2000. It is, however, a simple dataset with basic features including wind speed and gust data. The motivation of this study is to create an optimized model and thus previous research [8, 9] shows that supplemental atmospheric data can augment predictive performance significantly.

We use synoptic-scale pressure-level data over western North America to supplement the dataset. We decided to use the ERA5 [7] single pressure level atmospheric data which may be found at

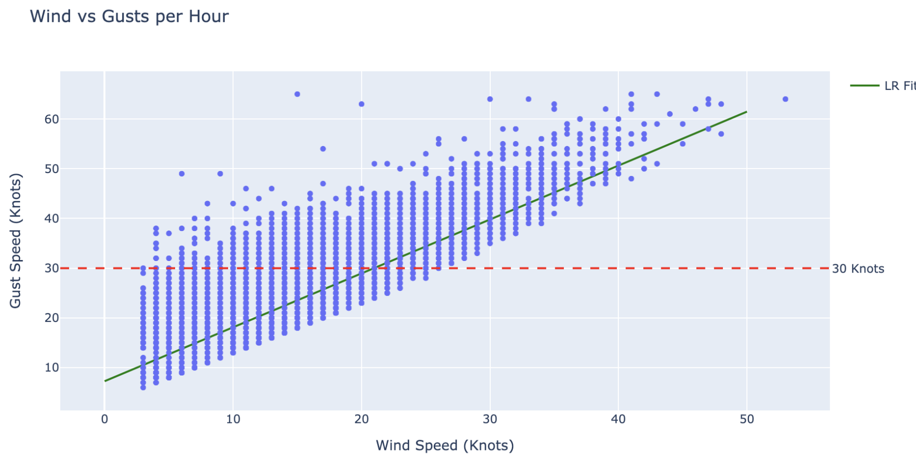


Figure 1: Linear regression fit of gust factor and wind speed.

`https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview`

The complete list of features is provided in the appendix. Although the ERA5 reanalysis product offers atmospheric variables at a spatial resolution of $0.25° \times 0.25°$, this study restricts input to full-degree grid points to reduce computational overhead. Despite this coarser spatial sampling, the inclusion of ERA5 data enables a regionally informed representation of Boulder's wind dynamics, augmented by synoptic-scale atmospheric predictors.

A notable limitation of the NCAR surface weather dataset is the presence of missing gust observations, with approximately 28% of gust values unavailable. To address this, we introduce a derived feature termed the *gust factor*, defined via a linear regression model:

$$\text{Gust Factor} = \alpha \cdot \text{WS} + \beta$$

where WS denotes wind speed, and α, β are regression coefficients estimated from the subset of the data for which both wind speed and gust measurements are available. This model achieves an R^2 score of 0.79, indicating a strong linear relationship. For time steps with missing gust values, the gust factor is used to impute estimates, while observed gust values are retained wherever present.

2.2 Reservoir Computing for Spatiotemporal Forecasting

Reservoir computing (RC) is a machine learning framework particularly well-suited for modeling nonlinear and chaotic dynamical systems. It consists of a fixed, high-dimensional recurrent neural network (the reservoir) that projects input signals into a rich dynamic space, followed by a trainable linear readout that maps the reservoir state to the desired output. This architecture enables efficient training, as only the readout weights are optimized, while the reservoir remains untrained.

In this study, RC is applied to forecast extreme wind events in Boulder, Colorado, using features derived from local meteorological observations. Each input to the reservoir consists of a delay-embedded representation of the time series, capturing both current and historical states of atmospheric variables.

The reservoir dynamics are governed by the update equation:

$$\mathbf{r}(t+1) = \tanh\left(W_{in}\mathbf{u}(t) + W_r\mathbf{r}(t)\right)$$

where $\mathbf{r}(t)$ is the reservoir state, $\mathbf{u}(t)$ is the input vector, W_{in} and W_r are fixed input and recurrent weight matrices, respectively. The output prediction is computed as:

$$\hat{y}(t) = W_{out}\mathbf{r}(t)$$

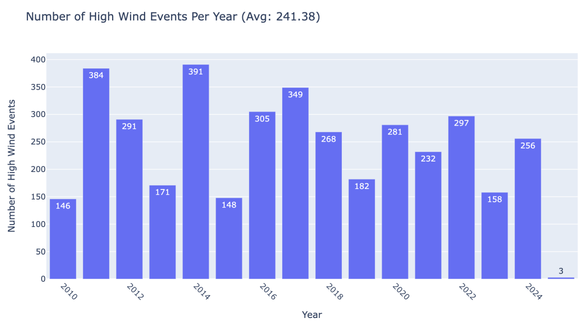
where W_{out} is learned via linear regression.

Reservoir computers were selected for their proven success in chaotic systems and efficient training compared to fully trainable recurrent networks [8]. Detailed performance metrics are discussed in 3.

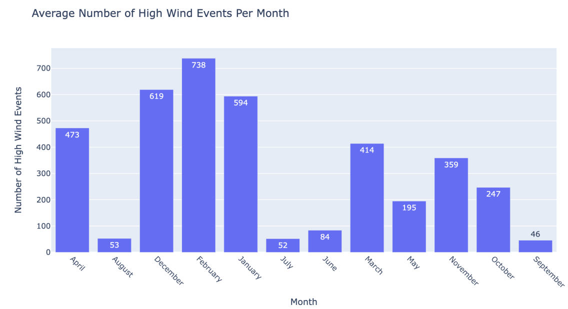
3 Results

This section presents the results of applying reservoir computing (RC) to the task of forecasting extreme wind events in Boulder, Colorado. We analyze both quantitative and qualitative aspects of model performance across a range of forecast horizons and feature configurations. The predictive task is framed as a binary classification problem: determining whether an extreme wind event—defined as a gust exceeding 35 knots—will occur within a specified future window.

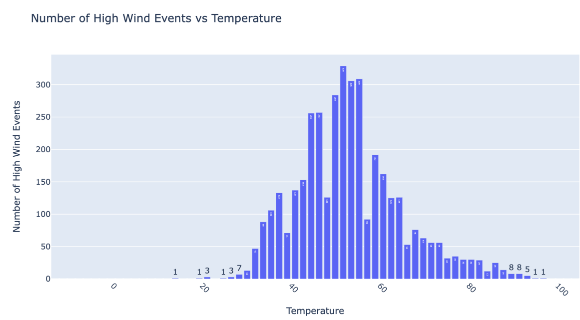
3.1 Temporal Distribution of Wind Events



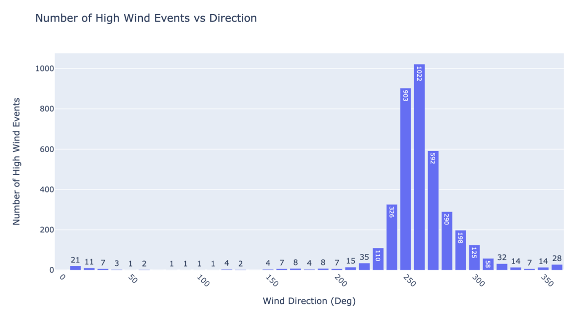
(a) Wind events per year.



(b) Wind events per month.



(c) Wind events per temperature.



(d) Wind events per flow direction.

Figure 2: Distributions of important features in high wind events.

From the dataset, we want to show when these wind events occur in Boulder. From the Figure 2, we see that these events happen most often in winter months, when conditions are warm and in the eastern direction. This will give us some insight to important factors that the model will capture.

3.2 Short-Term Predictive Performance

Reservoir computers showed strong performance in the short-term prediction regime. At a 12-hour forecast horizon (±30 minutes), the model achieved a sensitivity of 0.76 and a positive predictive value (PPV) of 0.83, resulting in a positive-class F1 score of 0.79. Negative-class F1 was 1.00, indicating perfect identification of non-events. These results suggest that the RC model captures sufficient memory and dynamics to anticipate the onset of high wind conditions over short intervals with relatively high confidence.

3.3 Forecast Horizon Sensitivity

As expected in chaotic systems, predictive performance deteriorated as the forecast horizon increased. For 24- to 72-hour windows, performance remained relatively high, with F1 scores above 0.80 for the positive class. However, forecasts at one week and one month displayed significant degradation, with F1 scores dropping below 0.20. This aligns with the known limits of predictability in chaotic systems, where errors in the initial condition grow exponentially over time.

Table 1: Reservoir Computing Performance Across Forecast Horizons

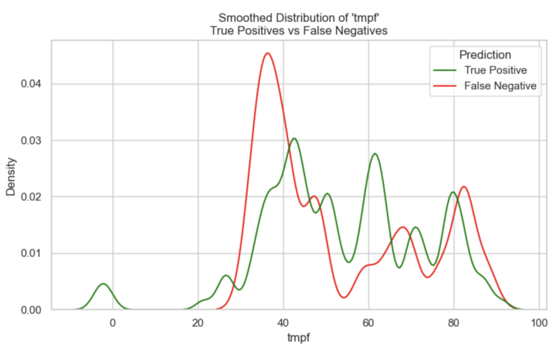
Time Horizon	Sensitivity	PPV	Positive F1	Negative F1
Random	0.51	0.14	0.22	0.64
12h	0.76	0.83	0.79	1.00
48h	0.86	0.81	0.83	0.99
72h	0.87	0.89	0.88	0.99
1 week	0.40	0.09	0.15	0.13
1 month	0.40	0.12	0.19	0.67

The rapid decline in PPV for long-range forecasts also suggests that while the model may still flag many events, its confidence in those predictions is less reliable. These findings highlight the natural limitations of forecasting within a chaotic framework and reinforce the utility of reservoir computing as a tool for short-term prediction.

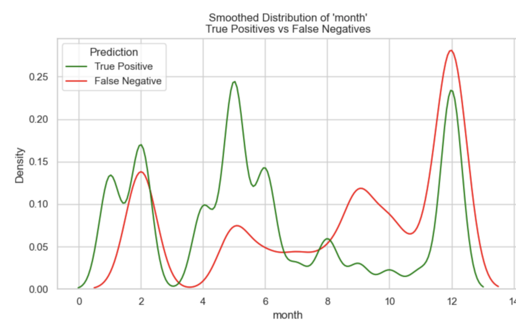
3.4 When Are Predictions Accurate?

To gain insight into when the reservoir model succeeds or fails, we analyze the distributions of selected atmospheric features for correctly and incorrectly predicted events. We compare true positives to false negatives across three key variables: temperature ('tmpf'), relative humidity ('relh'), and calendar month.

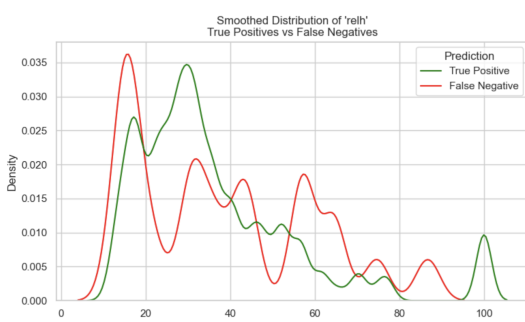
Figure 3 shows density estimates for each feature, comparing distributions.



(a) Temperature (`tmpf`): TPs tend to occur at higher temperatures than FNs.



(b) Month: FNs are more common in late fall and winter.



(c) Relative Humidity (`relh`): FNs skew toward higher humidity.

Figure 3: Smoothed distributions of true positives (green) and false negatives (red) for key meteorological variables. These plots help reveal contexts where prediction performance is strongest and weakest.

These comparisons reveal performance differences based on atmospheric context. For instance, the model tends to miss more events during colder months and at lower temperatures—potentially due to seasonal shifts in wind dynamics or underrepresentation in the training data. Similarly, a bias toward false negatives under high humidity conditions suggests different regimes that may require feature engineering or model specialization.

3.5 Practical Implications

From an operational standpoint, the model's ability to accurately predict extreme winds up to 72 hours in advance opens potential avenues for early-warning systems and infrastructure preparedness. The RC approach offers efficient training times and low computational overhead, making it feasible for deployment in low-latency, high-frequency forecasting environments.

However, its performance limitations at longer horizons and lack of interpretability present challenges. Future work may explore hybridizing reservoir computing with physical constraints or physics-informed neural networks (PINNs) to increase reliability and interpretability in operational settings.

4 Discussion

Reservoir computing proved to be an effective framework for forecasting extreme wind events in Boulder, particularly for short-term time horizons. However, the model's strengths must be considered against its interpretability, especially when considering its deployment in a scientific contexts.

One of the trade-offs in this study is between predictive power and model explainability. Reservoir computers excel at learning nonlinear temporal patterns without requiring explicit physical modeling, but they operate as a black-box. Unlike classical methods such as local linear forecasting or phase space reconstruction, reservoir models do not offer direct access to interpretable dynamical quantities, such as attractor geometry or Lyapunov exponents. While the RC may capture such structures within its high-dimensional reservoir state, these representations are not readily accessible for analysis. This raises an important question: can black-box models meaningfully recover or reflect the same low-dimensional dynamics captured by traditional embedding-based techniques?

The reservoir model exhibited strong performance for 12–72 hour predictions but declined sharply beyond that. This is consistent with the limitations of chaotic systems. But, while the reservoir trained successfully on the existing NCAR Mesa dataset, it remains to be tested whether its learned dynamics would generalize to other geographic locations or wind regimes without significant re-training.

Another factor influencing performance is the size of the prediction window. Experiments show that loosening the target window — for example, predicting whether an event occurs within ±4 hours rather than ±0.5 hours — significantly improves performance metrics. Wider windows naturally reduce temporal precision requirements, allowing the model to focus on identifying broader dynamics associated with wind onset, rather than pinpointing exact timing. This effect highlights the importance of task framings.

In terms of feature engineering, results suggest that incorporating thermodynamic and synoptic-scale context (e.g., temperature, pressure, and humidity) materially improves performance. These variables likely encode precursor conditions for downslope wind events, such as cold-air damming, lee-side pressure gradients, and boundary layer instability [3]. Their importance reinforces the notion that embedding physically meaningful signals can enhance machine learning performance and shows the importance of feature selection guided by meteorological expertise.

Several limitations of this study should be acknowledged. First, the reservoir model's assumptions — including fixed recurrent topology and static weights — may limit its expressiveness for more complex or multiscale systems. Secondly, there may be scale mismatches between the local station observations and the broader atmospheric drivers of extreme winds, which are non-local in space and time.

Ultimately, this study highlights the promise and limitations of reservoir computing in chaotic, spatiotemporal settings. It provides strong evidence for its use in short-term forecasting but also motivates further exploration into hybrid models, interpretable dynamical embeddings, and physically informed learning models.

5 Conclusion

This study shows the potential of RCs as a data-driven predictor for extreme wind events in Boulder, Colorado. Despite relying only on local atmospheric observations and minimal tuning, the reservoir model achieved high predictive performance for short-term horizons, particularly within 12 to 72 hours. These results affirm the capability of RCs to capture nonlinear temporal dependencies in spatiotemporal time series without the need for fully specified physical models.

A key finding is that model performance is strongly sensitive not only to the length of the forecast horizon, but also to the granularity of the prediction window. Additionally, the inclusion of thermodynamic and synoptic-scale features such as temperature, pressure, and humidity proved crucial for predictive accuracy, underscoring the value of physically informed feature selection even in black-box modeling pipelines.

Future work will explore hybrid modeling frameworks that combine the forecasting strength of reservoir computing with the physical insight offered by dynamical systems theory. This includes physics-informed reservoirs, and hybrid approaches that incorporate partial physical constraints or embedded knowledge of atmospheric dynamics.

In sum, this work supports the idea that reservoir computing can serve as an effective tool for forecasting in extreme weather systems.

References

[1] H. Poincaré. Les méthodes nouvelles de la mécanique céleste, Jan 1899.

[2] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, Mar 1963.

[3] Andrew E. Mercer, Michael B. Richman, Howard B. Bluestein, and John M. Brown. Statistical modeling of downslope windstorms in boulder, colorado. *Weather and Forecasting*, 23(6):1176–1194, Dec 2008.

[4] Alexander Marusov, Vsevolod Grabar, Yury Maximov, Nazar Sotiriadi, Alexander Bulkin, and Alexey Zaytsev. Long-term drought prediction using deep neural networks based on geospatial weather data, 2024.

[5] Daniel J. Gauthier, Erik Bollt, Aaron Griffith, and Wendson A. S. Barbosa. Next generation reservoir computing. *Nature Communications*, 12(1), September 2021.

[6] Federico Amato, Fabian Guignard, Sylvain Robert, and Mikhail Kanevski. A novel framework for spatio-temporal prediction of environmental data using deep learning. *Scientific Reports*, 10(1), December 2020.

[7] Nsf ncar mesa laboratory. `https://archive.eol.ucar.edu/cgi-bin/weather.cgi?site=ml&units=english`. Accessed: 2025-04-20.

[8] Daniel J. Gauthier, Erik Bollt, Aaron Griffith, and Wendson A. S. Barbosa. Next generation reservoir computing. *Nature Communications*, 12(1), September 2021.

[9] Kerry A. Emanuel. An air-sea interaction theory for tropical cyclones. *Journal of the Atmospheric Sciences*, 43(6):585–605, Mar 1986.

Appendices

A Code and Data

All code and dataset API requests are hosted on github.