Echo State Networks for Renewable Energy Forecasting

Samuel G. Dotson, Kathryn D. Huff

Dept. of Nuclear, Plasma and Radiological Engineering, University of Illinois at Urbana-Champaign
sgd2@illinois.edu

INTRODUCTION

Load following has given natural gas an economic edge over nuclear power because natural gas plants can follow grid demand and even shut off when renewable penetration makes the price of electricity go negative [1]. Some French nuclear plants have been retrofitted for load following capabilities to follow daily variations in electricity demand [2]. Unlike the United States, French nuclear power plants enjoy a majority share of the country's electric generation which makes daily variations predictable. Renewable energy has challenged the base load electricity production that nuclear provides in the United States by introducing grid demand variability that is much less predictable. This lack of predictability is primarily due to renewables' tight coupling with chaotic weather systems. Figure 1 shows that daily variations in electricity demand are reasonably predictable, with a usual evening minimum at 40 MW. When renewable energy is included in the mix, demand is much harder to follow.

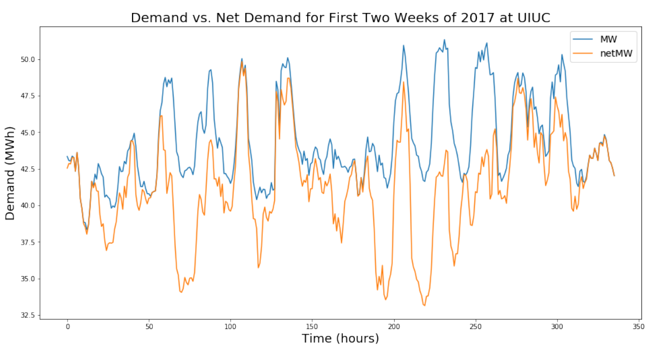

Fig. 1: Comparison between total demand and demand accounting for renewable energy. "NetMW" is the wind and solar production subtracted from the total.

Advanced reactor designs, like Molten Salt Reactors (MSRs), promise strong load following capabilities due to harder neutron spectra and faster Xe-135 burnup [3].

Unfortunately, the most mature MSR designs are at least a decade away from obtaining a commercial license in the United States. The climate crisis is too urgent to wait this long for nuclear power to become fully competitive with natural gas. Nuclear energy can be more economically feasible by relaxing the strong load following requirements with improved predictions of renewable energy production several hours or days in advance. If reactor operators knew in advance how much electricity will be produced by renewable energy they can slowly and accurately ramp reactor power to meet demand rather than operate continuously at full power and risk paying to export electricity. Thus improving nuclear energy's comptetitiveness against natural gas and strengthening nuclear's ability to couple with renewable energy. In this work we introduce Echo State Network (ESN) as a preferred method for time series forecasting of chaotic and stochastic systems like electricity production from renewable sources.

BACKGROUND

Variability has been the primary drawback for renewable energy sources like wind turbines, solar PV, and solar concentrators since their inception. This flaw has become more pronounced as renewable penetration on the electricity grid increased in recent years. Forecasting electricity production from renewable sources is therefore important for successful management of power systems [4]. Recent studies have applied artificial neural networks (ANNs), specifically multilayer perceptrons, to the task of net load forecasting [4, 5, 6]. These studies made short term forecasts of 4-6 hours. Forecasts at this time scale enable load following for natural gas plants, but challenges nuclear power plants. Nuclear plants need accurate forecasts further ahead to facilitate relaxed load following. This study will be the first to apply Echo State Networks (ESNs) to the task of net load prediction.

The University of Illinois at Urbana-Champaign is an ideal model system for this work because of its diverse energy mix. Previous work has been done to characterize this energy grid and optimize the size of a nuclear reactor [7]. Due to the degree of wind penetration, the University is sometimes forced to sell electricity back to the grid operator, MISO, at a loss because of overproduction from wind energy. Thus, a reliable prediction of electricity production from wind and other variable sources will reduce the likelihood of these events.

ESNs, a flavor of reservoir computing, are a modern machine learning algorithm that enables accurate short to medium term predictions. Pathak et. al used an ESN to predict the evolution of a chaotic system, a laminar flame front, up to seven Lyapunov times in the future [8, 9]. A Lyapunov time simply measures the timescale at which chaos makes initial predictions useless. The effect of chaos typically overwhelms conventional predictions after a single Lyapunov time, by definition. The Lyapunov time for a weather system is on the order of a few days but depends on the regional environment. ESNs have also been used to forecast multivariate time series [10]. Echo state networks are unique among neural networks in their ease of implementation and training speed. This is owed to its sparse network architecture [8, 9, 11]. However, their simplicity is balanced by the need for carefully chosen hyperparameters for the desired task [12]. Combining accurate demand and renewable energy predictions will enable an artificially intelligent reactor operator to adjust power in a relaxed manner.

METHODOLOGY

Echo State Networks

An "echo state network" (also called a "liquid state machine" [12]) is a type of recurrent neural network that uses a single layer of many neurons called a "reservoir". The reservoir has an adjacency matrix $\boldsymbol{A}$ that is

1. sparsely populated
2. connected by uniformly random weights centered at zero
3. has a large number of neurons

A reservoir computer also satisfies the *echo state property* [8, 13]. This property ensures that a system's state has a decaying influence on future states (like an echo of sound or ripples on water). This property is satisfied in most cases when the spectral radius (the absolute value of the greatest eigenvalue of $\boldsymbol{A}$)[13] is,

$$\rho(\boldsymbol{A}) < 1. \tag{1}$$

However, the echo state property can still be satisfied for a spectral radius greater than unity [12].

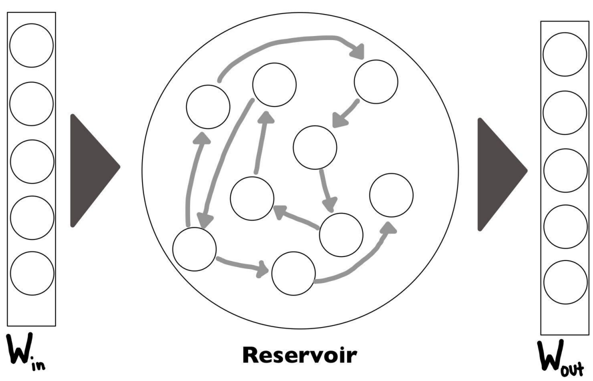

Fig. 2: A basic reservoir computer or echo state network. The connections in the reservoir are given by $\boldsymbol{A}$

Figure 2 gives a visual representation of a basic ESN. An input vector of length K is mapped to the reservoir layer by an input weight matrix $\boldsymbol{W_{in}}$. The state of the reservoir is mapped to an output layer of length N with an output weight matrix $\boldsymbol{W_{out}}$. An ESN does not require $K = N$, nor does it require $K \neq N$. In this work, the input vector is a function of time, $\boldsymbol{u}(t)$, and the output vector is the next state of the system, $\boldsymbol{u}_p(t + \Delta t)$. Ideally, the difference between the prediction, $\boldsymbol{u}_p$, and the actual, $\boldsymbol{u}_a$, is minimized. During training, the output weight matrix is trained through backpropagation using a loss function like cross entropy [8, 14].

Hyperparameter Search

Due to the architecture of ESNs, the weights and connections inside the reservoir do not need to be trained and, in our choice of implementation, cannot be. This dramatically reduces the training time because only the linear output layer needs to be trained. One drawback of this approach is its sensitivity to hyperparameters, which must be carefully chosen before running the network [8, 12, 13, 15]. Here, we perform grid searches to establish which combination of hyperparameters minimizes the mean squared error of the model,

$$MSE = \frac{1}{N}\sum_{i}^{N}(\hat{y} - y_i)^2 \tag{2}$$

where

$$\hat{y} = \text{the average value of the ouput.}$$

Model Prediction

As previously mentioned the weights of the output layer, $\boldsymbol{W_{out}}$ are trained through backpropagation by minimizing the error. When a trained model is given some initial state and then makes a prediction, $\boldsymbol{u}_p(t + \Delta t)$, this prediction becomes the initial state for the next prediction and so on. Even a good model, like in Pathak et. al [8], has some propagating error that deteriorates the prediction fidelity.

RESULTS

The hyperparameters of the ESN used in Figure 3 and Figure 4 were randomly assigned and therefore not optimized. In spite of this, preliminary predictions track reasonably well with grid demand. The current iteration shows a potentially misleading relationship between accuracy and training length. By inspection, Figure 4 is more accurate but the comparison is unfair because the two ESNs are predicting different time periods. We also conducted a single grid search for the optimal combination of spectral radius (ρ) and noise injection (for regularization of reservoir neurons), shown in Figure 5, following the recommendations from [12].

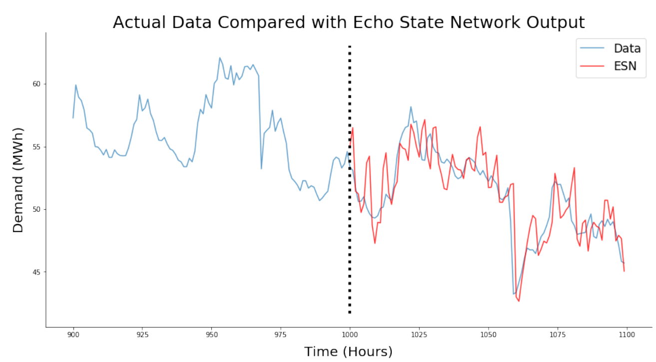

Fig. 3: A simple ESN with a prediction of 100 hours into the future after training on 1000 hours of historical data.

CONCLUSIONS

We have demonstrated that even a basic ESN can powerfully predict the evolution of dynamic, chaotic, systems as others have [8, 9, 10]. Future work will include:

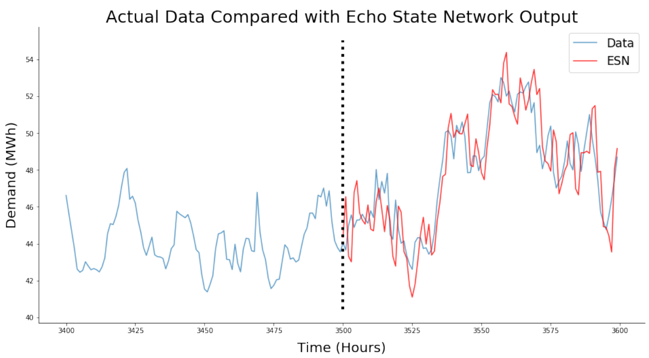

Fig. 4: A simple ESN with a prediction of 100 hours into the future. After training on 3500 hours of historical data.

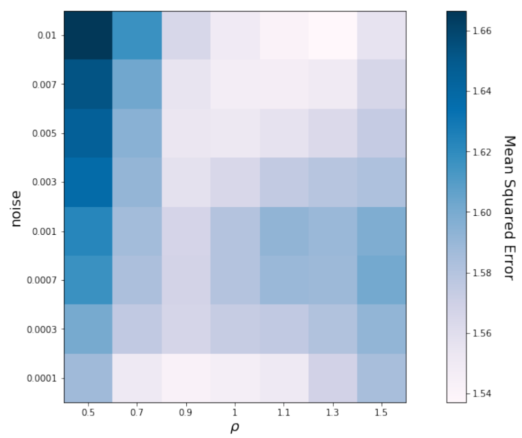

Fig. 5: A grid search over a range of spectral radii and noise levels. The optimal set minimizes the mean squared error.

1. Identifying ideal input vectors, whether a single value for net demand history will suffice or some combination of values (e.g. local weather and total demand) will improve predictive power.
2. Grid searches to tune hyperparameters.
3. Predicting generation from the UIUC solar farm and Railsplitter Wind Farm independently.

Fairer comparisons on the effectiveness of training length will be done by fixing the prediction time period for each ESN. Accurate predictions of chaotic systems, like wind energy production, will enable nuclear power plants to improve their economic feasibility through relaxed load following.

ACKNOWLEDGMENTS

This work was made possible with data provided by UIUC Facilities and Services, in particular, Morgan White, Mike Marquissee, and Mike Larson. Additionally, this work is supported by the Nuclear Regulatory Commission Fellowship Program. Prof. Huff is supported by the Nuclear Regulatory Commission Faculty Development Program (award NRC-HQ-84-14-G-0054 Program B), the Blue Waters sustained-petascale computing project supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois, the DOE ARPA-E MEITNER Program (award DE-AR0000983), and the DOE H2@Scale Program (Award Number: DE-EE0008832)

REFERENCES

1. J. H. KEPPLER, C. MARCANTONINI, O. N. E. AGENCY, and O. FOR ECONOMIC CO-OPERATION {AND} DEVELOPMENT, *Carbon pricing, power markets and the competitiveness of nuclear power*, Nuclear development, Nuclear Energy Agency, Organisation for Economic Co-operation and Development.
2. A. LOKHOV, "Technical and Economic Aspects of Load Following with Nuclear Power Plants," p. 53.
3. A. RYKHLEVSKII, D. O'GRADY, T. KOZLOWSKI, and K. D. HUFF, "The Impact of Xenon-135 on Load Following Transatomic Power Molten Salt Reactor," in "Transactions of the American Nuclear Society," American Nuclear Society.
4. P. KOBYLINSKI, M. WIERZBOWSKI, and K. PIOTROWSKI, "High-resolution net load forecasting for micro-neighbourhoods with high penetration of renewable energy sources," **117**, 105635.
5. S. DUTTA, Y. LI, A. VENKATARAMAN, L. M. COSTA, T. JIANG, R. PLANA, P. TORDJMAN, F. H. CHOO, C. F. FOO, and H. B. PUTTGEN, "Load and Renewable Energy Forecasting for a Microgrid using Persistence Technique," **143**, 617–622.
6. S. LEE, J.-H. RYU, B.-M. HODGE, and I.-B. LEE, "Development of a Neural Network-based Renewable Energy Forecasting Framework for Process Industries," in Z. KRAVANJA and M. BOGATAJ, editors, "Computer Aided Chemical Engineering," Elsevier, *26 European Symposium on Computer Aided Process Engineering*, vol. 38, pp. 1527–1532.
7. S. G. DOTSON, "Optimal Sizing of a Micro-reactor for Embedded Grid Systems," American Nuclear Society Annual Meeting 2020.
8. J. PATHAK, B. HUNT, M. GIRVAN, Z. LU, and E. OTT, "Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach," **120**, *2*, 024102, publisher: American Physical Society.
9. A. WIKNER, J. PATHAK, B. HUNT, M. GIRVAN, T. ARCOMANO, I. SZUNYOGH, A. POMERANCE, and E. OTT, "Combining machine learning with knowledge-based modeling for scalable forecasting and subgrid-scale closure of large, complex, spatiotemporal systems," **30**, *5*, 053111, publisher: American Institute of Physics.
10. F. M. BIANCHI, S. SCARDAPANE, S. LØKSE, and R. JENSSEN, "Reservoir computing approaches for representation and classification of multivariate time series," .
11. S. VANNITSEM, "Predictability of large-scale atmospheric motions: Lyapunov exponents and error dynamics," **27**, *3*, 032101.
12. M. LUKOŠEVIČIUS, "A Practical Guide to Applying Echo State Networks," in G. MONTAVON, G. B. ORR,

and K.-R. MÜLLER, editors, “Neural Networks: Tricks of the Trade: Second Edition,” Springer, Lecture Notes in Computer Science, pp. 659–686.

13. M. LUKOŠEVIČIUS and H. JAEGER, “Reservoir computing approaches to recurrent neural network training,” **3**, *3*, 127–149.
14. 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,” **126**, 191–217.
15. C. GALLICCHIO and A. MICHELI, “Deep Echo State Network (DeepESN): A Brief Survey,” .