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

Many nuclear power plants are at risk of shutting down due to a combination of policy decisions and financial strain [1]. Nuclear power competes with natural gas to fulfill the net demand after accounting for renewable energy sources. Load following gives 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 [2]. It is possible for existing nuclear plants to load follow somewhat and some French nuclear plants have been retrofitted to follow daily variations in electricity demand [3]. As penetration of solar and wind energy increases around the world, load following will depend less on electric demand profiles and more on renewable energy profiles. This shift makes load following with current nuclear power plants intractable [4]. Renewable energy challenges the base load electricity production that nuclear provides in the United States by introducing grid demand variability. Advanced reactor designs, like some Molten Salt Reactors (MSRs), promise strong load following capabilities due to active ^{135}Xe removal [5]. 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 [6].

Variability has been the primary drawback for renewable energy sources like wind turbines, solar PV, and solar concentrators, since their inception. This flaw worsens as renewable energy penetration increases [4]. Forecasting electricity production from renewable sources is therefore important for successful management of power systems [7]. Recent studies applied artificial neural networks (ANNs), specifically multilayer perceptrons, to the task of net load forecasting [7, 8, 9]. These studies made short term forecasts of 4-6 hours but nuclear plants need accurate forecasts further ahead to facilitate load following. This study will be the first to apply Echo State Networks (ESNs) to the task of net load prediction.

BACKGROUND

The University of Illinois at Urbana-Champaign is an ideal model system for this work because of its diverse energy mix. Previous work characterized this energy grid and optimized the size of a nuclear reactor for decarbonization [10]. 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. Figure 1 shows the total demand and net demand profiles at the University of Illinois at Urbana-Champaign (UIUC). Daily variations in electricity demand are reasonably predictable, with a usual evening minimum around 40 MW. Weekdays have peaks near 50MW and are distinguishable from weekends with peaks near 45MW. When renewable energy is included in the mix, regular changes from day/night and weekday/weekend are impossible to track by inspection. This lack of predictability is primarily due to renewables' tight coupling with chaotic weather systems. The chaotic nature of weather, and by extension renewable energy, indicates that there are dynamics that can be learned even without a physical model.

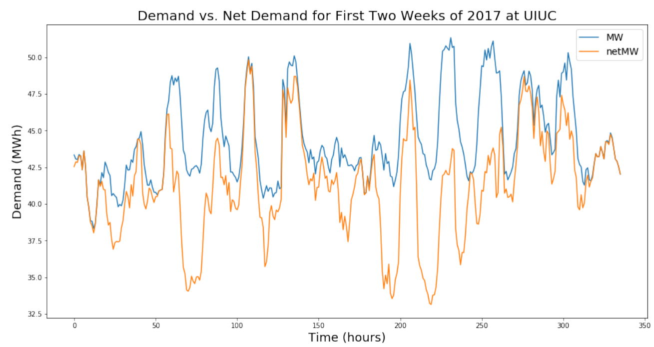

Fig. 1: Comparison between total demand and demand accounting for renewable energy. "netMW" is the total minus wind and solar.

ESNs, a flavor of reservoir computing, are a modern machine learning algorithm that can learn the dynamics of one chaotic system. Pathak et. al used an ESN to predict the evolution a laminar flame front up to seven Lyapunov times in the future [11, 12]. 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 [13]. ESNs are unique among neural networks in their ease of implementation and training speed. This is owed to its sparse network architecture [11, 12, 14]. However, their simplicity is balanced by the need for carefully chosen hyperparameters for the desired task [15]. Accurate renewable energy forecasts with ESNs could enable new load following strategies for nuclear power plants.

METHODOLOGY

Echo State Networks

An "echo state network" (also called a "liquid state machine" [15]) 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

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

A reservoir computer also satisfies the *echo state property* [11, 16]. 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 the absolute value of the greatest eigenvalue of $\boldsymbol{A}$ (the spectral radius, ρ) [16] is,

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

However, the echo state property can still be satisfied for a spectral radius greater than unity [15].

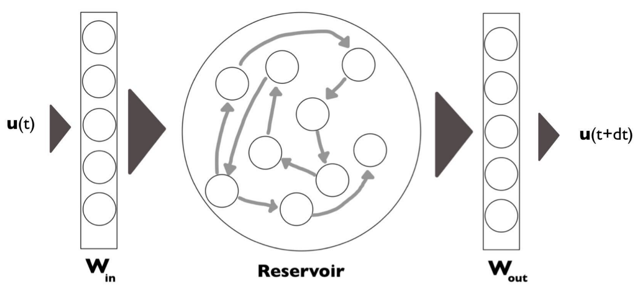

Fig. 2: A basic reservoir computer or echo state network. The connections in the reservoir are given by $\boldsymbol{A}$

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 $\boldsymbol{W_{out}}$. The output weight matrix is trained through backpropagation using a loss function like cross entropy [11, 17].

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. ESNs are capable of mapping an input vector of any size to an output vector of any size. For example, $\boldsymbol{u}(t)$ could be the total demand at t and $\boldsymbol{u}_p(t+\Delta t)$ could be the total demand at $t+\Delta t$. In that scenario the reservoir is a one-to-one map. Alternatively, $\boldsymbol{u}(t)$ could be several data points, temperature, air pressure, irradiance, wind speed, and total demand, at time t and $\boldsymbol{u}_p(t+\Delta t)$ could be the net demand at $t+\Delta t$. In this case the reservoir is a many-to-one map. Determining the combination of input data that leads to the most accurate predictions is a key area of research and an important part of this study [7, 13]. For the results shown below, $\boldsymbol{u}(t)$ is just the hourly demand.

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 [11, 15, 16, 18]. 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

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

Constructing the ESN

We constructed the ESN in our initial demonstration with the open source Python package `pyESN` [19]. The code is freely available on GitHub. Two models were trained with historical demand data from UIUC in 2017. Each model had a reservoir of 500 neurons. One was trained on 1000 hours of data and predicted the following 100 hours of demand in two hour increments. The other was trained on 3500 hours of data and, once again, predicted the following 100 hours.

A grid search informed our choice of hyperparameters for these models, shown in Figure 3. This search determined the optimal combination of spectral radius (ρ) and noise injection (for regularization of reservoir neurons), shown in Figure 3, following the recommendations from [15].

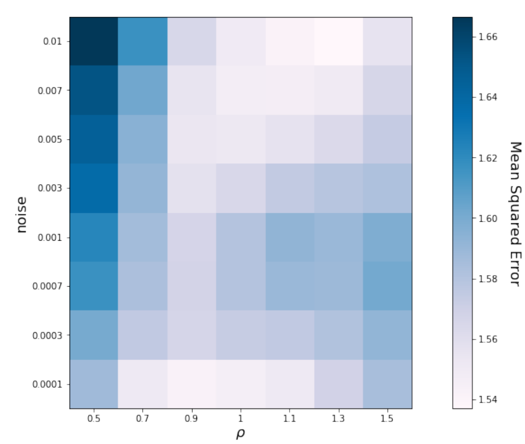

Fig. 3: A grid search over a range of spectral radii and noise levels. The optimal set minimizes the mean squared error.

RESULTS

The first model, shown in Figure 4, we trained with 1000 hours of historical data. The blue line is a snapshot of historical demand data. The dashed line indicates the end of the model training and the beginning of the model prediction. The red line is the prediction made by the ESN. By inspection, the ESN successfully predicted a general trend but had difficulty with accuracy at the desired hourly resolution.

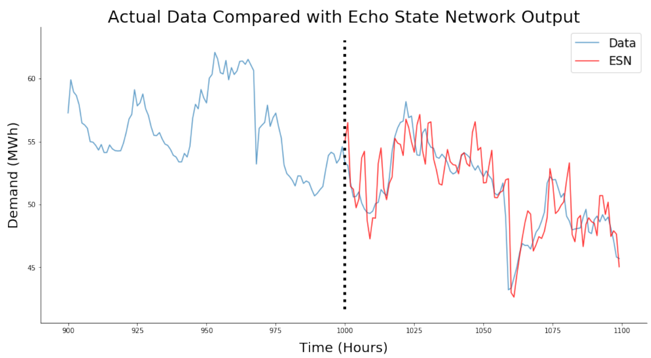

Fig. 4: A simple ESN with a prediction of 100 hours into the future after training on 1000 hours of historical data.

The second model, shown in Figure 5, we trained with the same dataset as before and extended the training to 3500 hours of historical data. This ESN made better predictions than the first model and is most likely due to longer training. The effect of training length on model accuracy will be explored in future work.

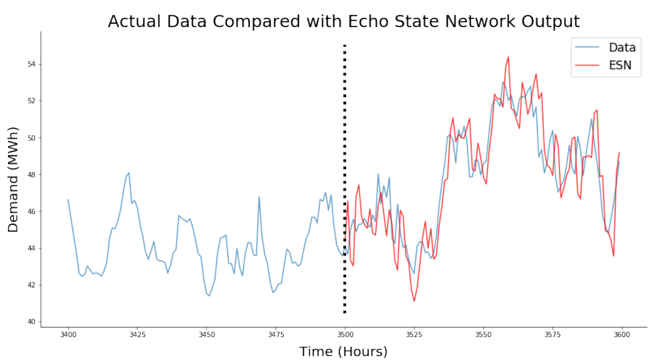

Fig. 5: A simple ESN with a prediction of 100 hours into the future. After training on 3500 hours of historical data.

Both models use training data with an hourly resolution. It is possible that data at the 15-minute or 30-minute timescale will improve prediction accuracy.

CONCLUSIONS

The ability of an ESN to predict dramatic changes in electricity demand may cause skepticism, especially given the simplicity of its implementation. It is important to note that a chaotic spatiotemporal system, like weather, is the result of complex dynamics and not randomness. We demonstrated that even a basic ESN can predict the evolution of a dynamic system, such as electricity demand. More work needs to be done to improve prediction accuracy for the purpose of load following with nuclear power plants. In the future we will use ESNs to forecast electricity production from solar and wind. We will also explore the impact of training length and network size on accuracy. Finally, we must determine an optimum set of input values for net demand prediction. These values might include temperature, wind speed, air pressure, and irradiance because of possible interactions among this data that could influence electricity production from renewable sources.

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. S. CLEMMER, J. RICHARDSON, S. SATTLER, and D. LOCHBAUM, "The Nuclear Power Dilemma," .
2. 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.
3. A. LOKHOV, "Technical and Economic Aspects of Load Following with Nuclear Power Plants," p. 53.
4. C. CANY, C. MANSILLA, G. MATHONNIÈRE, and P. DA COSTA, "Nuclear contribution to the penetration of variable renewable energy sources in a French decarbonised power mix," **150**, 544–555.
5. 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.
6. INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE, *Climate Change 2014 Mitigation of Climate Change: Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press.
7. P. KOBYLINSKI, M. WIERZBOWSKI, and K. PIOTROWSKI, "High-resolution net load forecasting for micro-neighbourhoods with high penetration of renewable energy sources," **117**, 105635.
8. 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.
9. 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.
10. S. G. DOTSON, "Optimal Sizing of a Micro-reactor for Embedded Grid Systems," American Nuclear Society An-

nual Meeting 2020.

11. 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.
12. 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.
13. F. M. BIANCHI, S. SCARDAPANE, S. LØKSE, and R. JENSSEN, “Reservoir computing approaches for representation and classification of multivariate time series,” .
14. S. VANNITSEM, “Predictability of large-scale atmospheric motions: Lyapunov exponents and error dynamics,” **27**, *3*, 032101.
15. 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.
16. M. LUKOŠEVIČIUS and H. JAEGER, “Reservoir computing approaches to recurrent neural network training,” **3**, *3*, 127–149.
17. 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.
18. C. GALLICCHIO and A. MICHELI, “Deep Echo State Network (DeepESN): A Brief Survey,” .
19. C. KORNDÖRFER, “pyESN,” .