Martin Rodriguez

P.I. Christof Teuscher, Ph.D

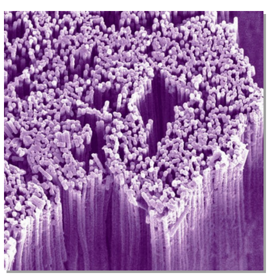
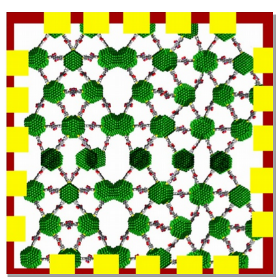
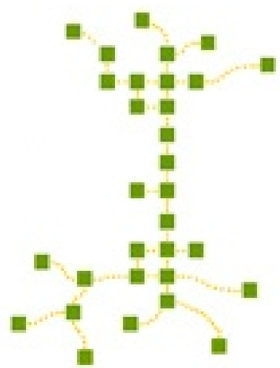
Exploring Training Techniques in Reservoir Computing

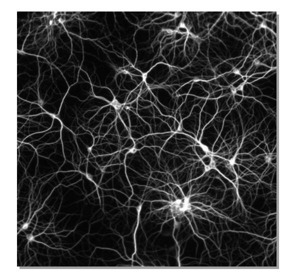
REU Symposium, 12 August 2016

Portland State University

Department of Electrical and Computer Engineering (ECE)

www.teuscher-lab.com | mtr@pdx.edu

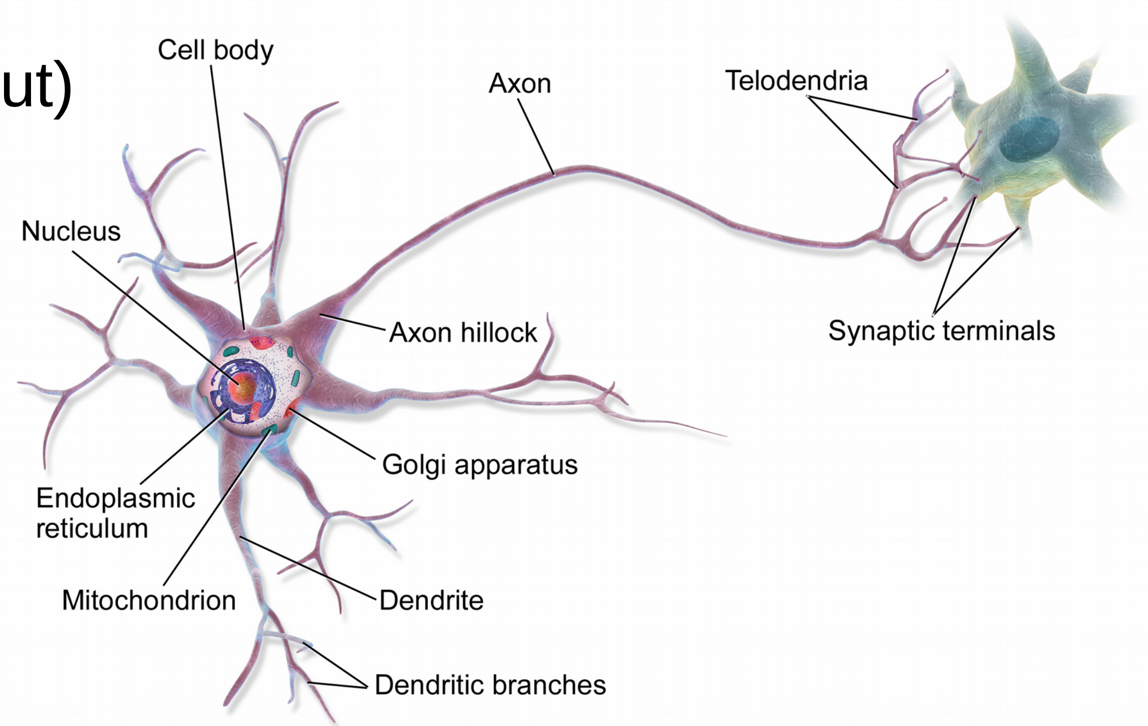


teuscher.:Lab
teuscher-lab.com
Emerging Computing Models and Technologies

Portland State
UNIVERSITY

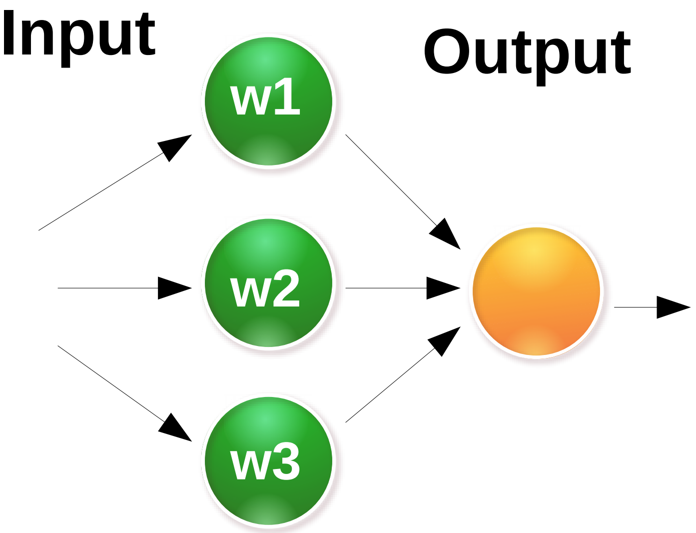
Biological Inspiration for Neural Networks

- Soma/cell body (input)
- Dendrites (inputs)
- Axon (output)
- Information is propagated through activation of neuron and resulting action potential down axon



What Makes Up a Neural Network (NN)?

- Perceptrons
  - Inputs must be linearly separable
  - Linear classifier (binary)
  - Weights are used multiplicatively to modify connections
  - Multilayer perceptrons can form larger networks
    - Are universal function approximators
  - Multi-class perceptrons can handle more categories
- Sigmoid neurons / other models
- Dynamic networks
  - Feedforward / Recurrent
  - Can be made of perceptrons / other models
  - Hopfield networks
  - Boltzmann machines
  - Echo state networks

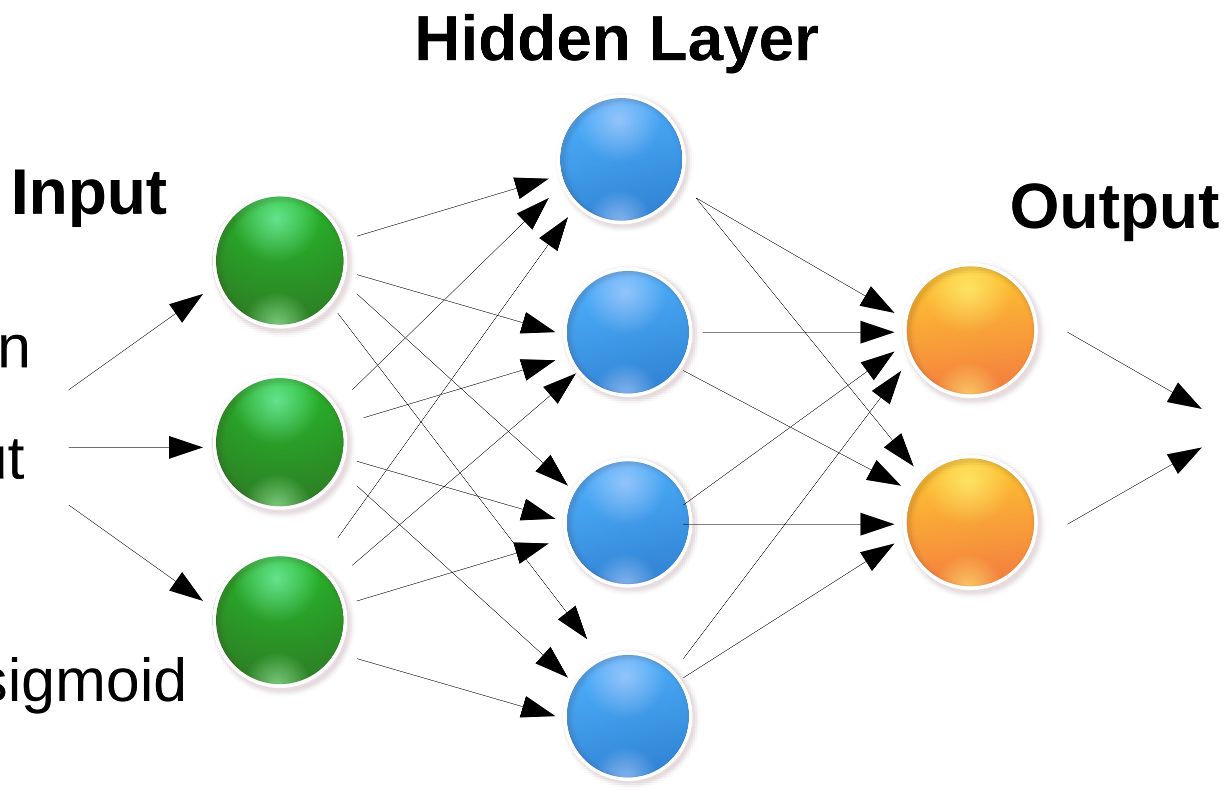


Characteristics of Networks

- **Architecture**
  - Density / sparsity
  - Arrangement of nodes
    - Circular, feedforward
- **System Dynamics**
  - Weights of connections between nodes
- **Learning**
  - How the weights change with time

Composition of a Feed-forward NN (Multi-Layer Perceptron)

- Information moves in one direction (forward)
- Input layer
- Hidden layer
- Output layer
- Activation function determines output of each node, typically tanh or sigmoid



Recurrent NNs

- Can be arranged more “creatively”
- Architecture produces cyclical properties
- Retain memories of past inputs, but have trouble “seeing” long patterns
- RNNs are harder to train

Limitations of Networks

- Calculations of new weights between nodes in hidden layers can be computationally expensive
- No one network is suited to all tasks
- Different training algorithms are needed for different tasks and it can be a challenge to find the optimal algorithm
- Optimizing initial parameters involves guess work

Solutions?

- Long-Short Term Memory
- Varying architectures (circular, sparsely connected, etc.)
- Hierarchical networks
- Better training algorithms?

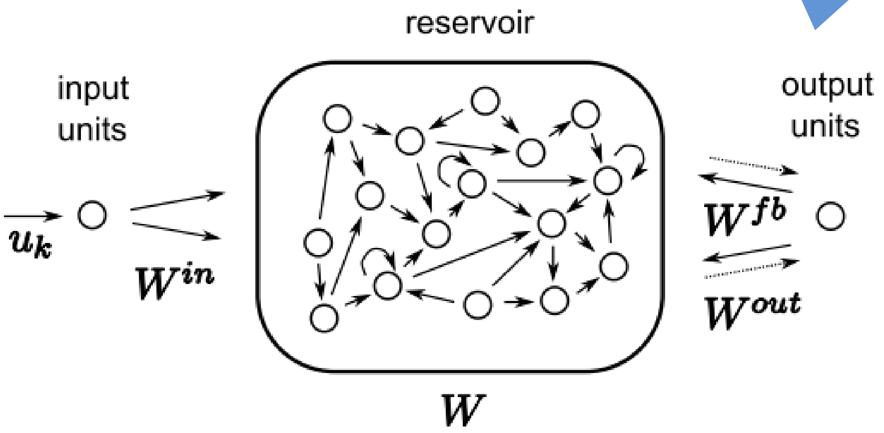
What is a Reservoir Computer?

- Also called an echo state network (ESN)
- Random, sparse architecture of nodes and weights
  - Internal weights are due to dynamical properties of system itself
- We train the output layer, which can produce specific functions based on its adapted weights and the fixed weights of the hidden layers
- Even a bucket of water can be a reservoir computer [3]

[3] Fernando, Chrisantha; Sojakka, Sampsa; Of Series Lecture Notes In Computer Science, ISBN (2005), "Pattern Recognition in a Bucket", In Advances in Artificial Life: 978–3

Training Methods (Simplified)

- Ordinary Linear Regression (OLR)
- Least squares
- Recursive Least Squares (RLS)



[13]

[13] Yildiz, I. B., Jaeger, H., & Kiebel, S. J. (2012). Re-visiting the echo state property. Neural Networks, 35, 1–9. http://doi.org/10.1016/j.neunet.2012.07.005

Benchmarks

- **Nonlinear Autoregressive Moving Average (NARMA-10)**

$$y(t) = 0.3\,y(t-1) + 0.05\,y(t-1)\sum_{i=1}^{10} y(t-i) + 1.5\,u(t-10)\,y(t-1) + 0.1$$ [2]

- **Mackey-Glass Time Series**

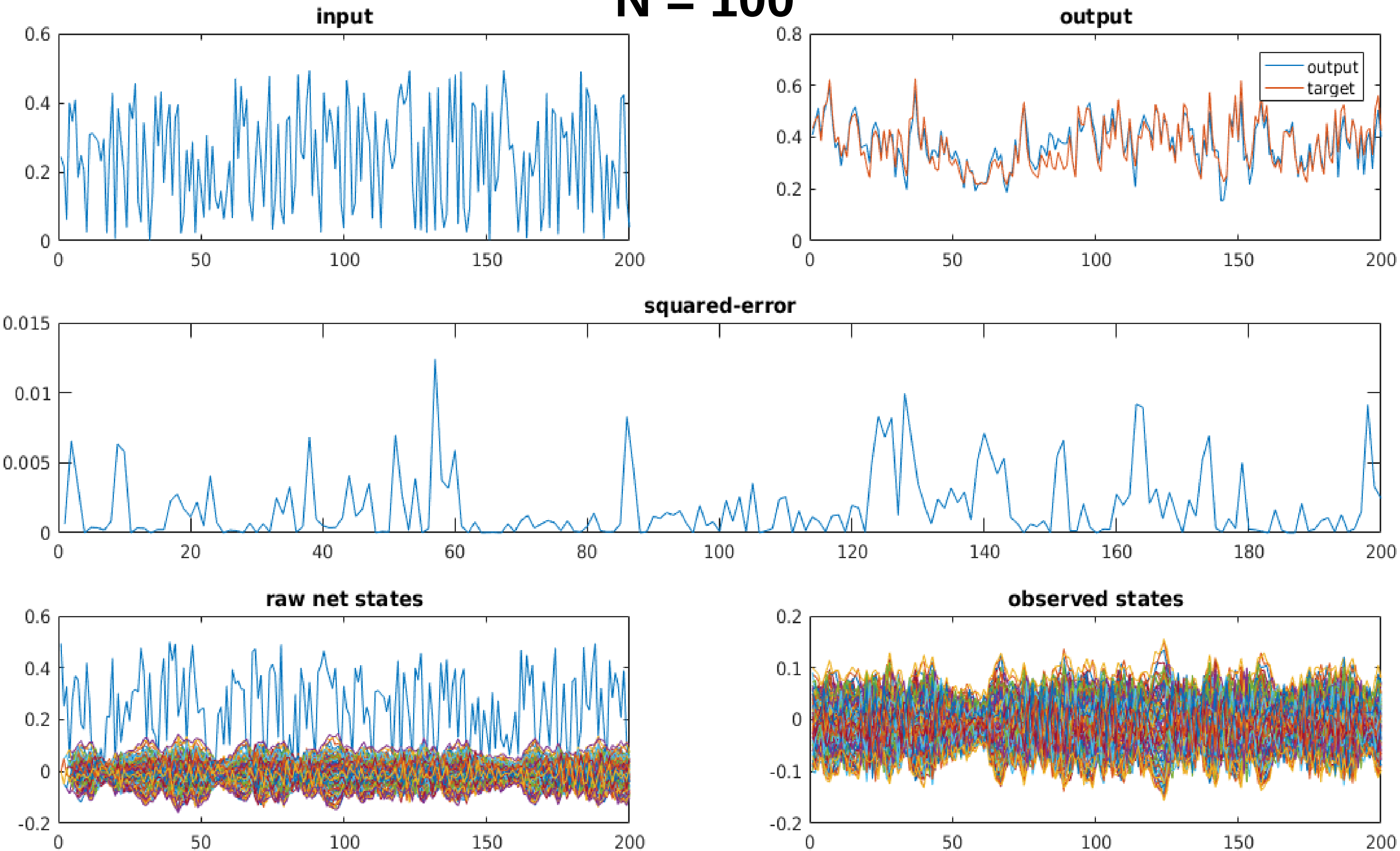
$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1+x(t-\tau)^{10}} - bx(t)$$ [2]

- **Sine wave prediction**

[2] Čerňanský, M., & Tiňo, P. (2008). Predictive Modeling with Echo State Networks. 18th International Conference on Artificial Neural Networks, (1), 778–787.

Results of OLR on NARMA-10 Task

N = 100

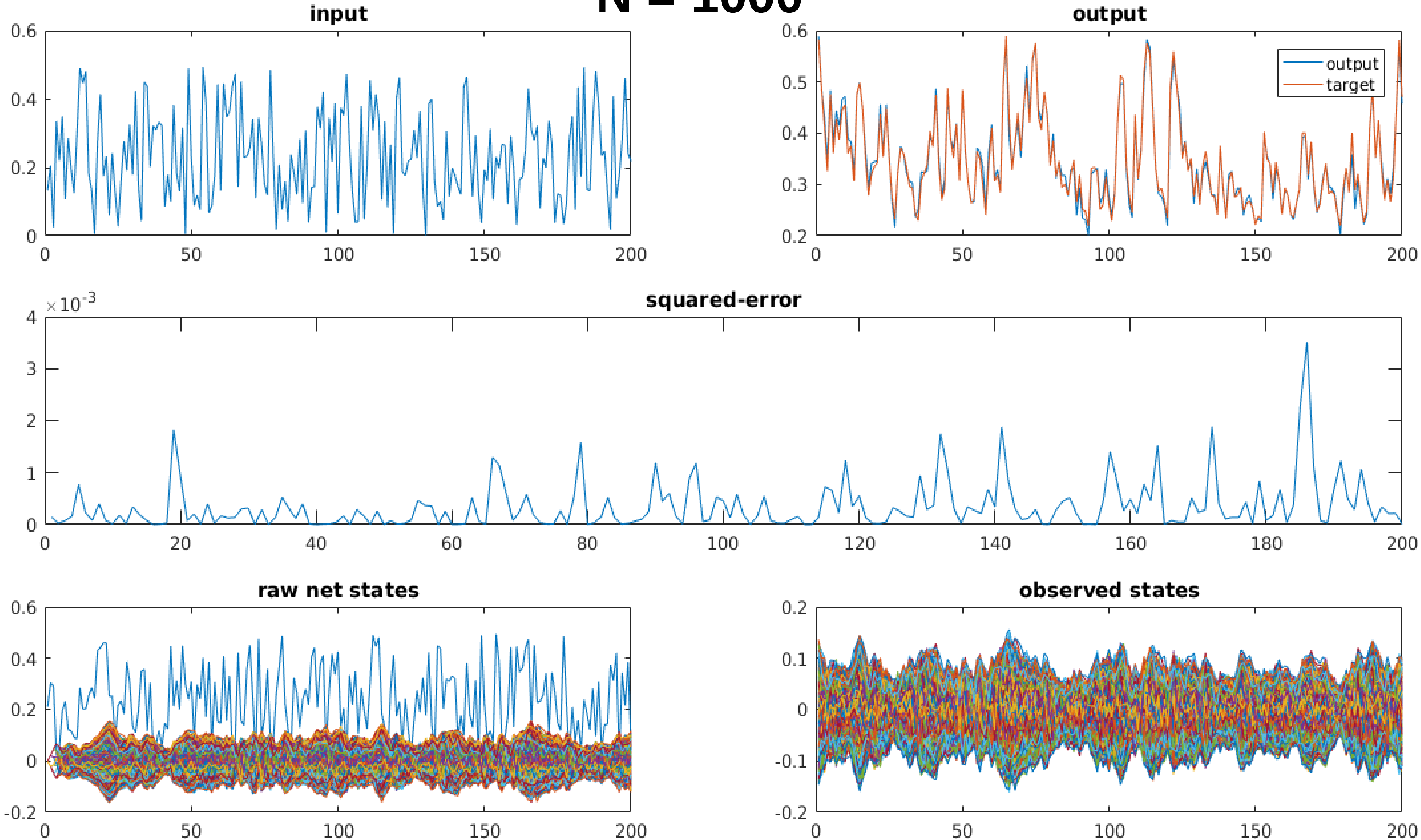


ESN framework by A. Goudarzi

NMSE = 0.1283

Results of OLR on NARMA-10 Task

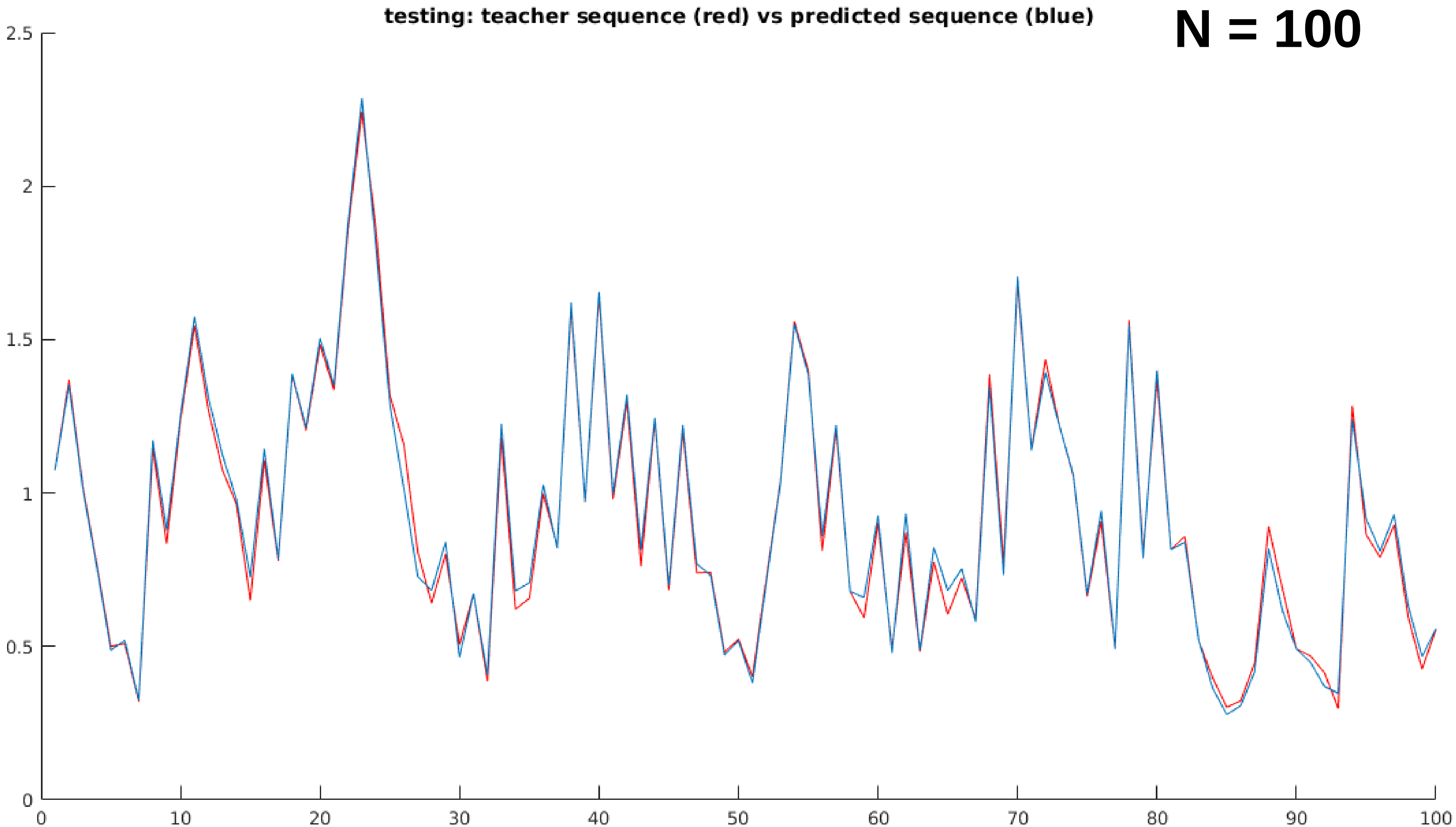
N = 1000



ESN framework by A. Goudarzi

NMSE = 0.0274

Results of OLR (batch) NARMA-10



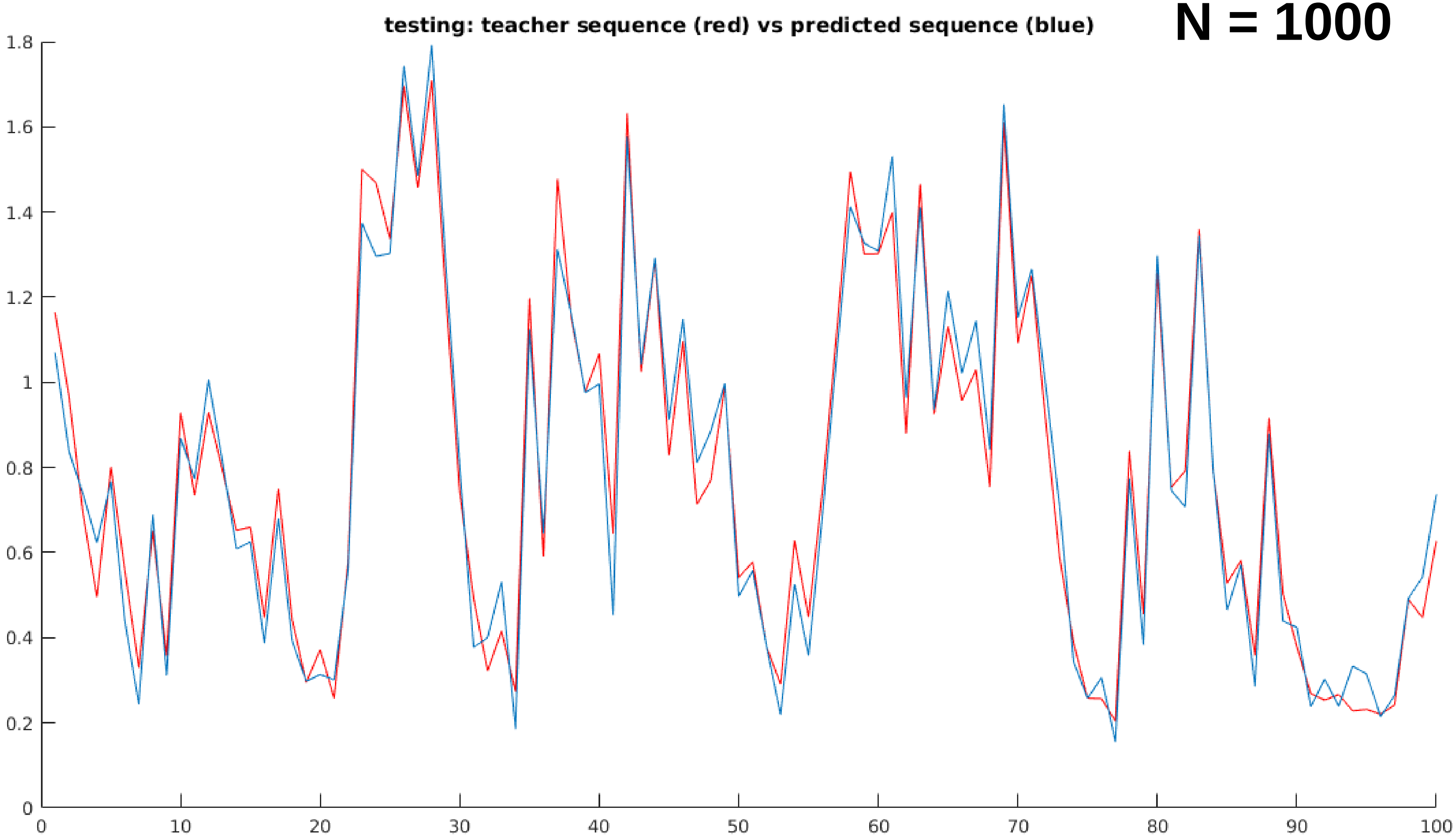
N = 100

ESN toolbox by H. Jaeger

NRMSE = 0.11899

Results of OLR (batch) NARMA-10

N = 1000

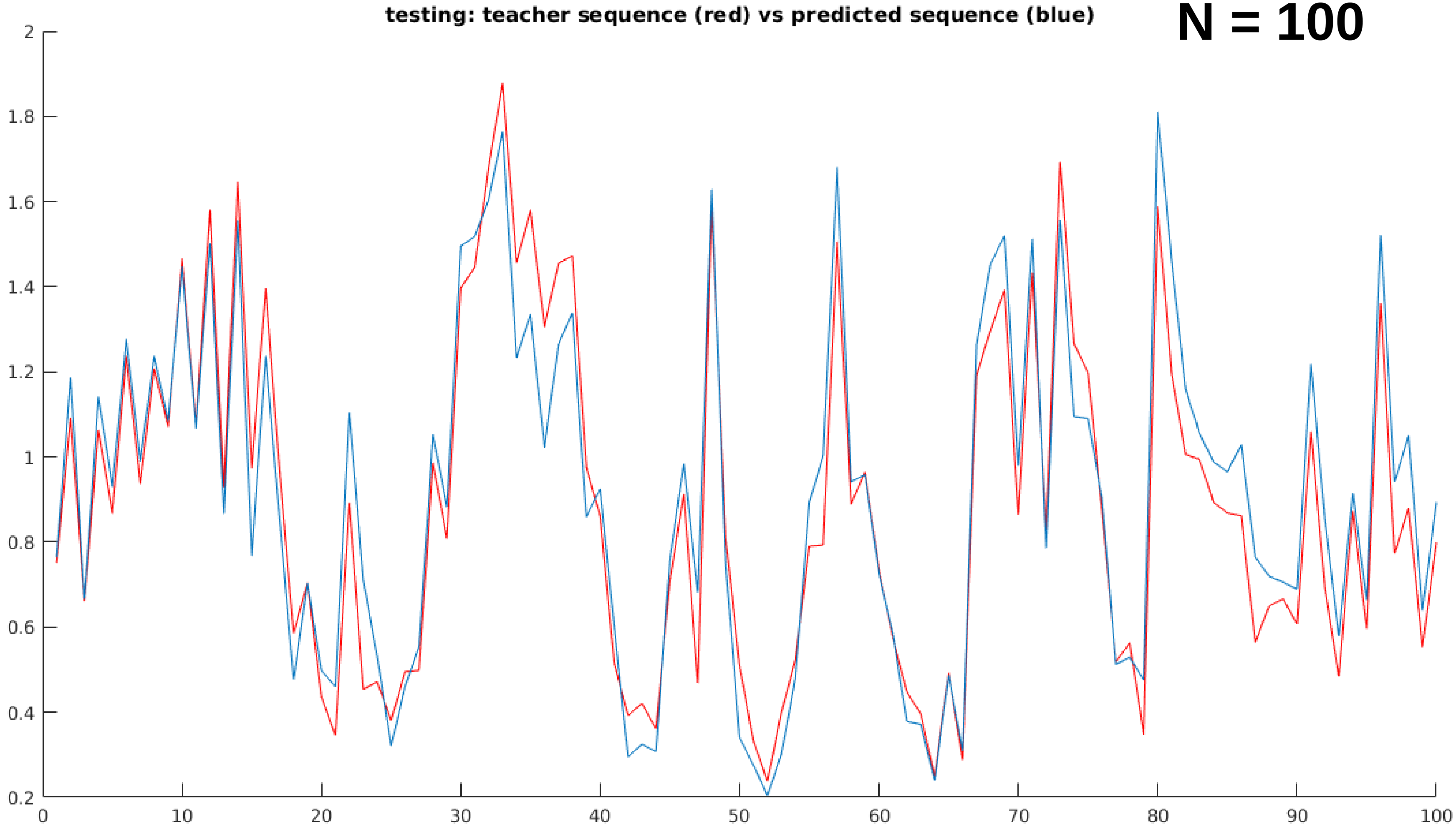


ESN toolbox by H. Jaeger

NRMSE = 0.16854

Results of RLS (online) NARMA-10

N = 100

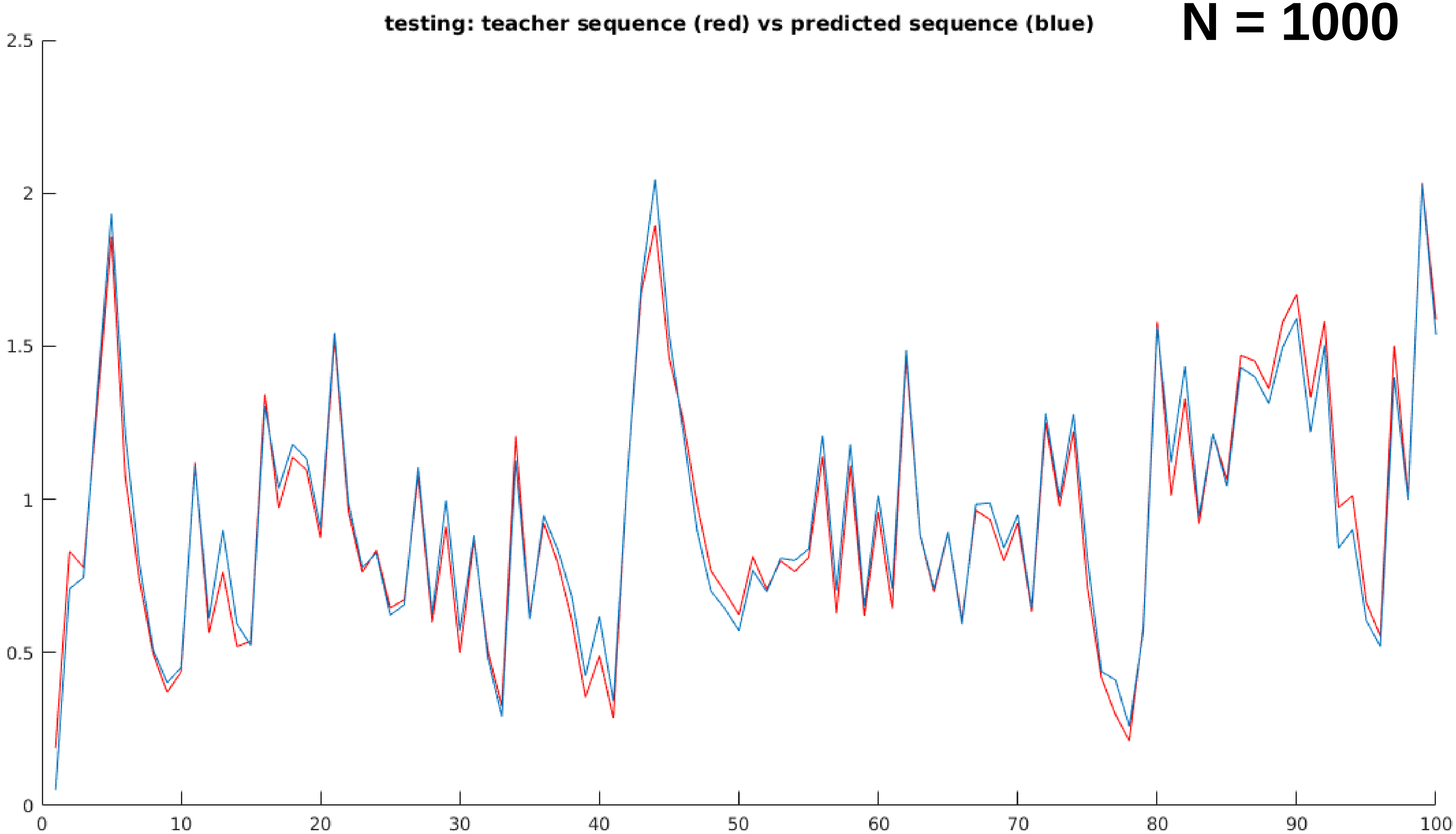


ESN toolbox by H. Jaeger

NRMSE = 0.29867

Results of RLS (online) NARMA-10

N = 1000



NRMSE = 0.16362

ESN toolbox by H. Jaeger

Conclusions

- RLS performs well for implementation in online learning schemes
- More benchmarks need to be tested to evaluate strengths and limitations of RLS
- Initial parameters may be further optimized to improve performance
- Next steps:
  - NARMA-30
  - Mackey-Glass time-series
  - Need to compare RLS to OLR in online contexts

References

[1] Atiya, A. F., & Parlos, A. G. (2000). New results on recurrent network training: unifying the algorithms and accelerating convergence. IEEE Transactions on Neural Networks, 11(3), 697–709. http://doi.org/10.1109/72.846741

[2] Čerňanský, M., & Tiňo, P. (2008). Predictive Modeling with Echo State Networks. 18th International Conference on Artificial Neural Networks, (1), 778–787.

[3] Fernando, C., & Sojakka, S. (2003). Pattern Recognition in a Bucket. Advances in Artificial Life, 588–597. http://doi.org/10.1007/978-3-540-39432-7_63

[4] Ferreira, A. A., Ludermir, T. B., & De Aquino, R. R. B. (2013). An approach to reservoir computing design and training. Expert Systems with Applications, 40(10), 4172–4182. http://doi.org/10.1016/j.eswa.2013.01.029

[5] Goudarzi, A., Lakin, M. R., & Stefanovic, D. (2013). DNA reservoir computing: A novel molecular computing approach. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8141 LNCS, 76–89. Neural and Evolutionary Computing; Adaptation and Self-Organizing Systems; Chaotic Dynamics; Biological Physics. http://doi.org/10.1007/978-3-319-01928-4_6

[6] Goudarzi, A., Lakin, M. R., & Stefanovic, D. (2014). Reservoir computing approach to robust computation using unreliable nanoscale networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8553 LNCS, 164–176. http://doi.org/10.1007/978-3-319-08123-6_14

[7] Haykin, S. O. (2008). Neural Networks and Learning Machines. http://doi.org/978-0131471399

[8] Jaeger, H. (2002). Adaptive Nonlinear System Identification with Echo State Networks. Advances in Neural Information Processing Systems (NIPS), 593–600. Retrieved from http://books.nips.cc/nips15.html

[9] Jaeger, H. (2013). ESNTutorialRev, 2, 1–46.

[10] Liu, X., Cui, H., Zhou, T., & Chen, J. (2012). Performance evaluation of new echo state networks based on complex network. The Journal of China Universities of Posts and Telecommunications, 19(1), 87–93. http://doi.org/10.1016/S1005-8885(11)60232-X

[11] Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. Computer Science Review, 3(3), 127–149. http://doi.org/10.1016/j.cosrev.2009.03.005

[12] Rodan, A., & Tino, P. (2011). Minimum complexity echo state network. IEEE Transactions on Neural Networks, 22(1), 131–144. http://doi.org/10.1109/TNN.2010.2089641

[13] Yildiz, I. B., Jaeger, H., & Kiebel, S. J. (2012). Re-visiting the echo state property. Neural Networks, 35, 1–9. http://doi.org/10.1016/j.neunet.2012.07.005

Acknowledgments

This project was funded by a grant from the National Science Foundation (NSF) through the Research Experience for Undergraduates program.

Thank you Dr. Christof Teuscher, Alireza Goudarzi, Dr. Jun Jiao, Audrey Siefert, and Dr. Erik Sanchez!