SCADA Data Modeling for Wind Turbine Gearbox Failure Detection Using Machine Learning and Big Data Technologies

Rafael Orozco, Georgia Tech
Shawn Sheng, Caleb Phillips, and Lindy Williams,
National Renewable Energy Laboratory

AIRES Workshop, ORNL
Oak Ridge, Tennessee
January 23, 2020

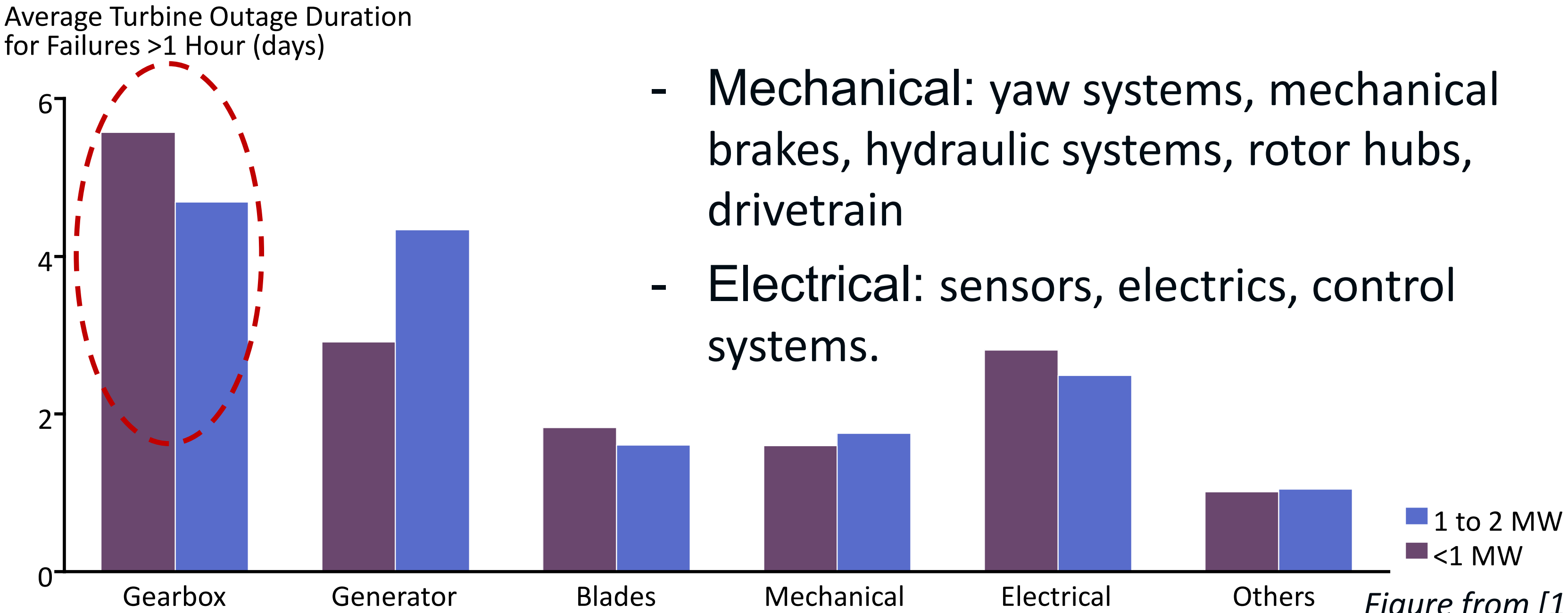
Outline

- Introduction
- Methodology
- Results
- Summary and Future Outlook

U.S. Department of Energy 1.5-MW turbine, NREL 17702

Wind Turbine Gearbox Reliability Challenge

- Downtime caused by premature component/subsystem failures, led by gearboxes, challenges the wind industry and results in an increased cost of energy for wind power.



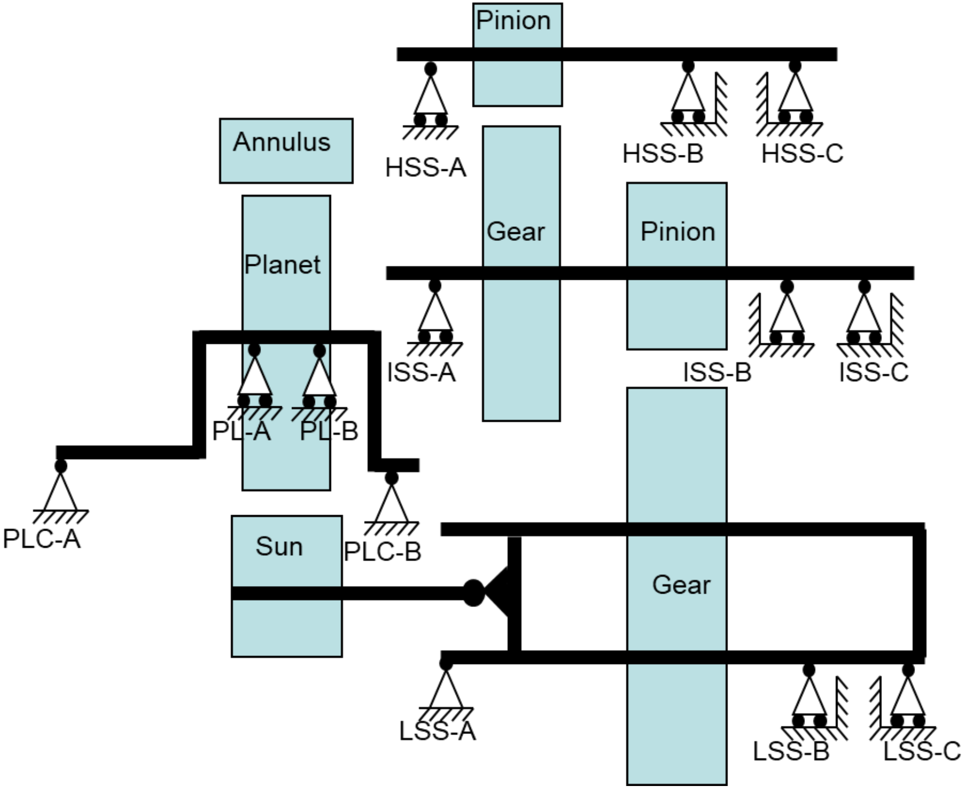
- Mechanical: yaw systems, mechanical brakes, hydraulic systems, rotor hubs, drivetrain
- Electrical: sensors, electrics, control systems.

Figure from [1]

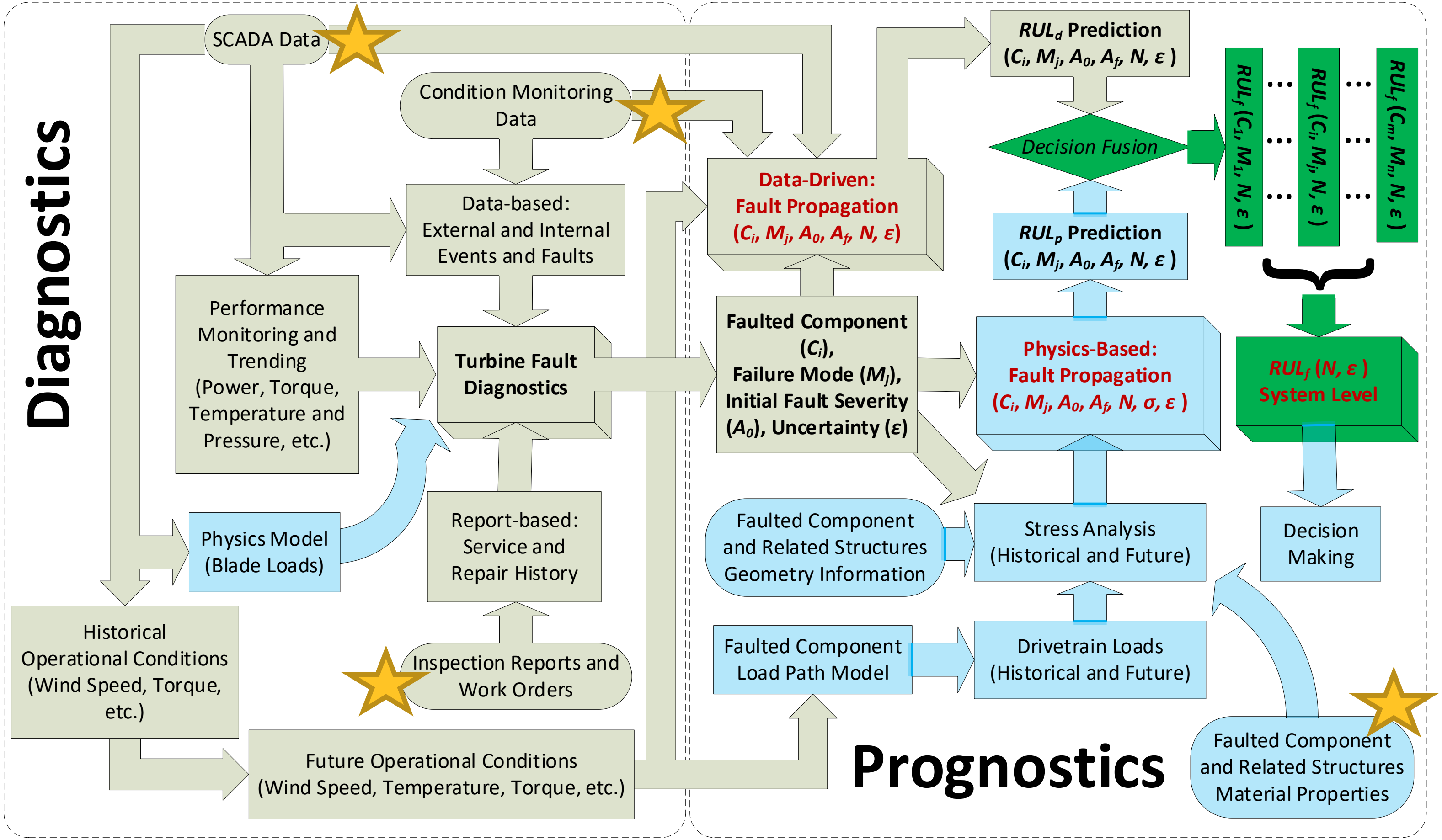
A Typical Wind Turbine Gearbox

DOE 1.5-MW Turbine Gearbox and Main Shaft

Photo by Dennis Schroeder, NREL 49418

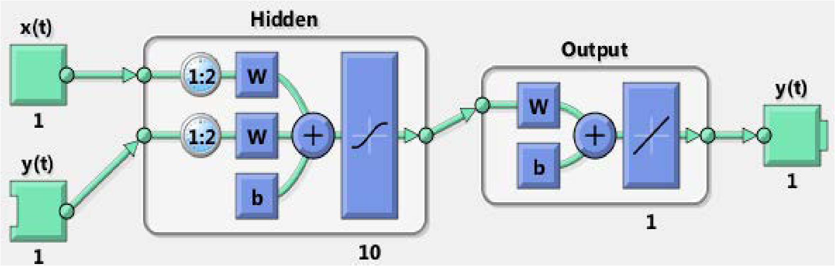


Operational Reliability: A Prognostics and Health Management for Wind Framework [2]



SCADA Time Series Data Modeling for Wind Turbine Fault Diagnosis [3]

- **Typical approach**: ***develop*** *models for normal operational states based on historical data, and use these models to* ***compare*** *model outputs with observed data to* ***identify*** *abnormal behaviors*
  - Most used ***relatively small data sets without much automation,*** and the developed baseline models are ***not scalable or hard to generalize***.
- **Objective**: ***to develop and demonstrate <u>a scalable and autonomous diagnostics modeling methodology</u> for wind turbine gearbox failures based on SCADA data using machine learning and big data technologies***



NREL's Eagle computer

Photo by Dennis Schroeder, NREL 53840

Methodology

SCADA Data

- Continuous Reliability Enhancement for Wind (CREW) data [4]
  - Compiled by Sandia National Laboratories and shared with NREL
  - Aiming to characterize reliability performance issues and identify opportunities for improvements
- Data summary: 614 turbines from 7 different plants, providing 388 years of turbine data, plant rated at about 970 MW, turbine rated between 1.5 and 2.05 MW, and native resolution varies from 2 to 8 seconds.

Wind Plant ID	Turbine Numbers	Turbine Days	Plant-Rated Power (MW)	Turbine-Rated Power (MW)	Native Resolution
1	41	20,902	61.5	1.5	5
2	147	22,653	207.5	1.5, 1.6	2
3	69	25,863	103.5	1.5	2
4	102	39,863	153	1.5	2
5	53	5,291	108.65	2.05	7–8
6	66	20,832	132	2	2
7	136	6,259	204	1.5	5–6
All Plants	614	141,666	970.15	1.5, 1.6, 2, 2.05	2–8

Data Preprocessing

- Down-sampling
  - Take moving average of all time-series SCADA features
  - Select sampling resolution of 300 measurements/window, corresponding to 10 minutes of data at 2 seconds native resolution, but larger (e.g., 20 minutes) for lower native resolution data (e.g., 4 seconds).
- Data cleaning
  - Data not standardized among different wind plants
  - Plant 6 did not have the temperature readings needed for the proposed modeling methodology and was excluded from results reporting.

Modeling of Normal Temperature Behavior

- Models investigated:
  - Linear regression: $T_m(i) = \alpha_0 + \alpha_1 * T_e(i) + \alpha_2 * P(i)$
  - Multivariate polynomial regression: $T_m(i) = \alpha_0 + \alpha_1 * T_e^2(i) + \alpha_2 * P\left(i\right) * T_e\left(i\right) + \alpha_3 * P^2\left(i\right)$
  - Random forest [5]: 16 trees used because increased number of trees did not show substantial increase in performance
  - Neural network [6]: a multilayer perceptron, 2-18-1 structure.

- Libraries used:
  - Scikit-learn [7]: machine learning library for Python
  - MLLib [8]: Apache Spark's scalable machine-learning library.

Proposed Automatic Failure Detection Process

1. Train a model mapping ambient temperature and power output to component temperature
2. Evaluate residual based on the trained model outputs and observed component temperature
3. Identify the data points corresponding to the component temperature being abnormally high as a spike
4. Check those data points identified in Step 3 to determine whether they are followed by a <u>turbine shutdown</u> (defined as power output being zero while wind resource is still appropriate to generate power)
5. Flag those data points followed by a turbine shutdown as an occurrence of the component failure (caution: may not be the same as defined by the end users and can be fine-tuned in future work)
6. Repeat Steps 1 to 5 to flag all failures for a turbine
7. Repeat Steps 1 to 6 to perform a plant- or fleet-wide failure analysis.

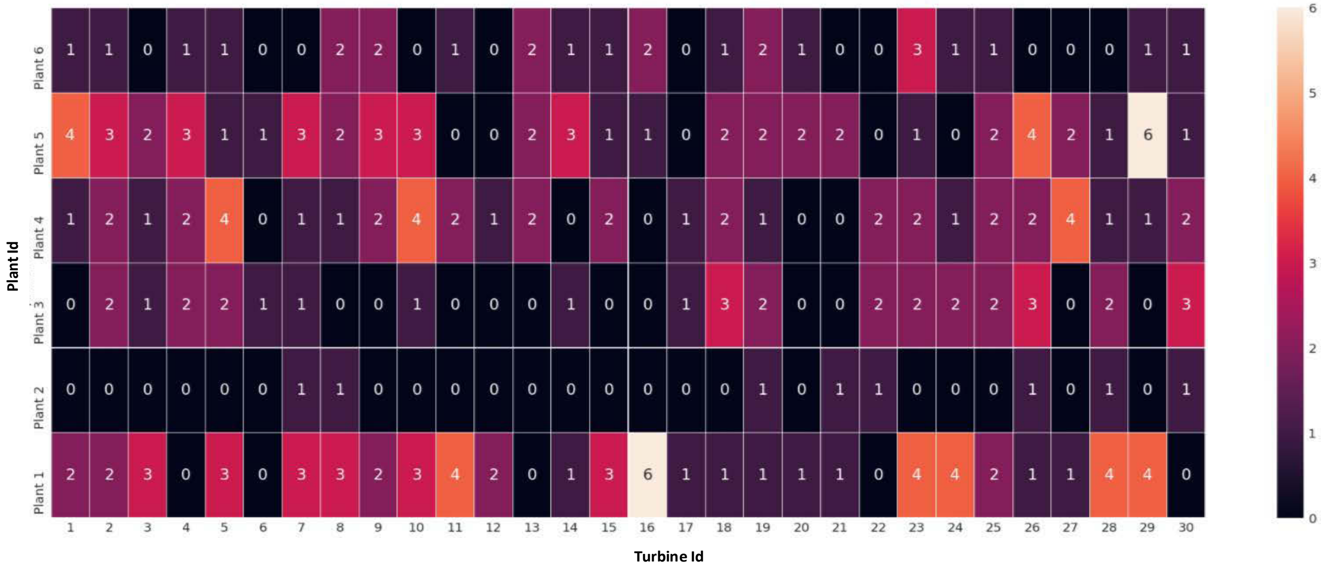
Scalability: Processing Data at Scale

- CREW data set is about 1 TB and needs a scalable solution
- The main library used was Spark, allowing algorithms to be run on a Spark cluster
- All codes were run on the NREL Sparkplug cluster environment, which implements the Hortonworks Data Platform [9] within Openstack [10]
- The specifications on the cluster are: 5 cores per executor and 20 executor nodes with 23 GB of memory each.

Results

Failure-Detection Statistics for All Plants

- A heat map generated showing number of failures occurred to the first 30 turbines of the studied six plants
  - Each cell represents one turbine labeled with the number of failures detected; each row represents the first 30 turbines from a plant
  - Turbines 16 and 29 in plants 1 and 5 showed the highest number of failures



Summary and Future Outlook

Summary and Future Work

- A SCADA time-series data-modeling method for wind turbine gearbox failure detection was developed and demonstrated by using machine learning (ML) algorithms and big data technologies
  - Automated and scalable, which will be increasingly important as operational data from wind plants become large and complex [11].
- Ongoing and Future Work:
  - Validate the introduced modeling methodology against actual field failure records
  - Expand to other major turbine components: generators
  - Explore more ML algorithms: light/extreme gradient boosting
  - Integrate with physics-domain modeling methods.

Challenges and Opportunities

- Data access: Reference datasets for model development and validation
- Data preprocessing: Not much standardized and consumes a lot of time
- Model developments: Dynamic operation conditions and regimes
  - Various turbine makers and models
  - Different components and diverse/complex failure modes
  - Offshore wind and wave condition modeling and forecasting
  - Model transferability: problem and application driven?
  - Black box: how can AI help improve understanding of physics
- Paradigm: reactive, condition-based maintenance, prognostics and health management to digital twin
- Impacts: cost effectiveness
  - Simple neural networks vs Deep learning neural networks
  - Edge computing, cloud storage, and parallel processing
  - Minimize maintenance trips, component replacement frequencies, downtime, and costs.

Siemens 3.6-MW Offshore Wind Turbines

Photo by Axel Schmidt, NREL 27824

References

1. Groupe Reaction Inc. and K. Chamberlain. 2015. *WEU Onshore Asset Optimization & Reliability Benchmarking Report 2015.* Technical Report. London, United Kingdom.
2. Sheng, S., and Y. Guo. 2019. “A Prognostics and Health Management Framework for Wind.” Paper No. GT2019-91533. Proceedings of ASME Turbo Expo 2019. June 17–21. Phoenix, Arizona: 9 pp. https://doi.org/10.1115/GT2019-91533.
3. Orozco, R., S. Sheng, and C. Phillips. 2018. “Diagnostic Models for Wind Turbine Gearbox Components Using SCADA Time Series Data.” Preprint. 11 pp. https://www.nrel.gov/docs/fy18osti/71166.pdf.
4. Peters, V., A. Ogilvie, and C. Bond. 2012. “CREW Database: Wind Plant Reliability Benchmark.” Sandia National Laboratories. Albuquerque, New Mexico.
5. Liaw, A., and M. Wiener. 2002. “Classification and Regression by Random Forest.” *R. News,* pp. 18–22.
6. Reder, M. D., E. Gonzalez, and J. Melero. 2016. “Wind Turbine Failures—Tackling Current Problems in Failure Data Analysis.” *Journal of Physics: Conference Series,* vol. 753, no. 7.
7. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al. 2011. “Scikit-Learn: Machine Learning in Python.” *Journal of Machine Learning Research,* vol. 12, pp. 2825–2830.
8. Apache Software Foundation. 2017. “MLLib—Apache Spark's Scalable Machine Learning Library.” http://spark.apache.org/mllib/.
9. Hortonworks. “Open Source Solutions for Data-at-Rest.” http://hortonworks.com.
10. Omar, S., M. Aissaoui, and M. Eleuldj. 2012. “Openstack: toward an open-source solution for cloud computing.” *International Journal of Computer Applications,* vol. 55, no. 3.
11. Yang, W., P. J. Tavner, C. J. Crabtree, Y. Feng, and Y. Qiu. 2014. “Wind turbine condition monitoring: technical and commercial challenges.” *Wind Energy,* vol. 17, no. 5, pp. 673–693.

Thank you!

shawn.sheng@nrel.gov

303-384-7106

www.nrel.gov

NREL/PR-5000-76056

Data sharing by Sandia National Laboratories and wind plant owner operator partners is acknowledged.

This work was supported in part by the U.S. Department of Energy (DOE) Office of Science and Office of Workforce Development for Teachers and Scientists under the Science Undergraduate Laboratory Internships Program. This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the DOE Office of Energy Efficiency and Renewable Energy Wind Energy Technologies Office. The views expressed in the presentation do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

NREL
Transforming ENERGY