

GENERALIZABLE TEMPERATURE NOWCASTING WITH PHYSICS-CONSTRAINED RNNs FOR PREDICTIVE MAINTENANCE OF WIND TURBINE COMPONENTS

Johannes Exenberger

TU Graz

johannes.exenberger@tugraz.at

Matteo Di Salvo

Sirius Energy Automation

disalvo@sirius-ea.com

Thomas Hirsch

TU Graz

Franz Wotawa

TU Graz

Gerald Schweiger

TU Graz

ABSTRACT

Machine learning plays an important role in the operation of current wind energy production systems. One central application is predictive maintenance to increase efficiency and lower electricity costs by reducing downtimes. Integrating physics-based knowledge in neural networks to enforce their physical plausibility is a promising method to improve current approaches, but incomplete system information often impedes their application in real world scenarios. We describe a simple and efficient way for physics-constrained deep learning-based predictive maintenance for wind turbine gearbox bearings with partial system knowledge. The approach is based on temperature nowcasting constrained by physics, where unknown system coefficients are treated as learnable neural network parameters. Results show improved generalization performance to unseen environments compared to a baseline neural network, which is especially important in low data scenarios often encountered in real-world applications.

1 INTRODUCTION

Machine learning (ML) methods are a crucial part of the wind energy sector (Marugán et al., 2018) and help improving operational efficiency of wind power plants to make electricity prices competitive with currently dominant fossil-based energy sources (Shafiee & Sørensen, 2019). Predictive maintenance is applied to prevent component failures and increase maintenance efficiency. Models of component deterioration processes estimate when maintenance should be performed before damages occur, resulting in shorter downtimes. For large wind farms and those with restricted access such as offshore plants, maintenance is costly and time consuming (Carroll et al., 2017). Optimized maintenance scheduling is thus highly desired, with estimated reductions in the Levelized Cost of Electricity of up to 30% for onshore wind farms (Costa et al., 2021). Especially gearbox faults, mainly caused by bearing failure, can lead to long downtimes (Hahn et al., 2007; Reder et al., 2016; Wang et al., 2020). The core of predictive maintenance is a surrogate model - either data-driven or physics-based - that describes the system in normal conditions. Deviations between this baseline and current measurements triggers an alarm. The system dynamics are often only partially known, making physics-based models difficult to adapt to each scenario when system-specific measurements are not available, while data driven methods don't guarantee physical plausibility. The integration of physical laws in ML models known as physics-informed ML (PIML) (Karniadakis et al., 2021) allows to combine both approaches. In this paper, we describe a neural network-based approach for predictive maintenance of a wind turbine generator (WTG) gearbox bearing constrained by physics without the need for precise measurements of component coefficients, which are treated as trainable parameters^{1,2}. By performing nowcasting, i.e. predicting the current state rather than a future state, (1) the inference process does not depend on wind speed forecasting which introduces additional

¹The code is available at: github.com/jxnb/pcrnn-wtg

²The dataset is available upon request to the authors.

uncertainties, and (2) external constraints such as for curtailment periods do not affect the model. This approach shows competitive prediction performance and improved generalization capabilities in comparison to a physics-based model and a baseline neural network. The approach is not limited to WTGs, but applicable to a wide range of systems where partially known physics-based knowledge can be integrated in the modelling process.

2 MODEL DESCRIPTION

We apply a common procedure of ML-based predictive maintenance: a surrogate model trained on observations represents the expected system behavior in normal conditions. Over the system’s lifetime, the deviations between estimated and true behavior are expected to increase with growing wear of the components, providing information about the condition of the system and potential need for maintenance. Our approach aims to circumvent common problems in ML-based predictive maintenance for WTGs. First, data on system faults is usually scarce - WTGs have availability rates in the order of 98% (Yang et al., 2013); component faults are rare and occur mostly at the end of a component lifetime period. This limits the applicability of fault detection methods. Second, ML models leveraging physics often require knowledge of specific component coefficients (e.g. Yucesan & Viana, 2022). This information is often not available or unreliable, as components like the gearbox vary even between WTGs of the same model. Treating these unknown coefficients as trainable parameters allows to incorporate physics into the model when they are not or only partially known, simplifying the adaptation of this approach to other applications. Third, wind speed is the main driver of bearing temperatures in WTGs, making component temperature prediction mainly a wind speed forecasting problem. Local wind speeds are very difficult to predict also for short time horizons of several minutes due to complex local turbulence patterns. As we focus on nowcasting, i.e. the estimation of the current bearing temperature state, the current rotor speed as the main driver is known. This allows us to avoid the large factor of uncertainty associated with wind speed forecasts and increases prediction performance during periods where wind speed and power production are uncorrelated, such as curtailment for wind farms due to, for example, grid loads constraints or bird protection.

2.1 PHYSICS MODEL

The physics-based description of a change in bearing temperature T^b between two time steps is given by the following heat transfer ODE equation for transient-state, lumped systems based on Zhang (2014); Cambron et al. (2017):

$$C_p \frac{dT^b}{dt} \approx R^{-1}(T_{t+1}^a - T_t^b) + \mu\omega_{t+1} + \alpha P_{t+1} \quad (1)$$

where C_p [W K⁻¹] is the nacelle heat capacity, $R^{-1}(T_{t+1}^a - T_t^b)$ is the system’s thermal conductivity with the resistance coefficient R [K W⁻¹], bearing temperature T_t^b [K] and ambient temperature T_{t+1}^a [K]. $\mu\omega_{t+1}$ is the heat produced by friction in the rotating gearbox shaft with friction coefficient μ [Ws²rad⁻²] and rotational frequency ω_{t+1} [rad s⁻¹]. αP_{t+1} is the amount of energy degradation as heat per unit of power P [KW] produced. The coefficients C_p , R , μ and α are unknown and depend on the complexity of the system and materials design and have to be approximated by the model. Based on Equation 1, the equation solved by the neural network with trainable network parameters $\lambda_1, \lambda_2, \lambda_3$ can be formulated as:

$$\Delta T_{t-1}^b \approx \lambda_1(T_t^a - T_{t-1}^b) + \lambda_2\omega_t + \lambda_3 P_t \quad (2)$$

2.2 PHYSICS-CONSTRAINED RECURRENT NEURAL NETWORK

Given the physics-based model in Equation 1, the observed state of a WTG at time t is described by the vector $\mathbf{x}_t = (T_t^a, \omega_t, P_t, T_{t-1}^b)$. The objective of the neural network is the prediction of the bearing temperature at current time t , \hat{T}_t^b . We employ a simple LSTM-based (Hochreiter & Schmidhuber, 1997) recurrent neural network (RNN) architecture applicable to a wide range of time-series problems. Physical constraints are imposed by integrating information based on the idea of physics-informed neural networks (PINNs) (Raissi et al., 2019), resulting in a physics-constrained recurrent neural network (PC-RNN). The model consists of a single-layer LSTM followed by a

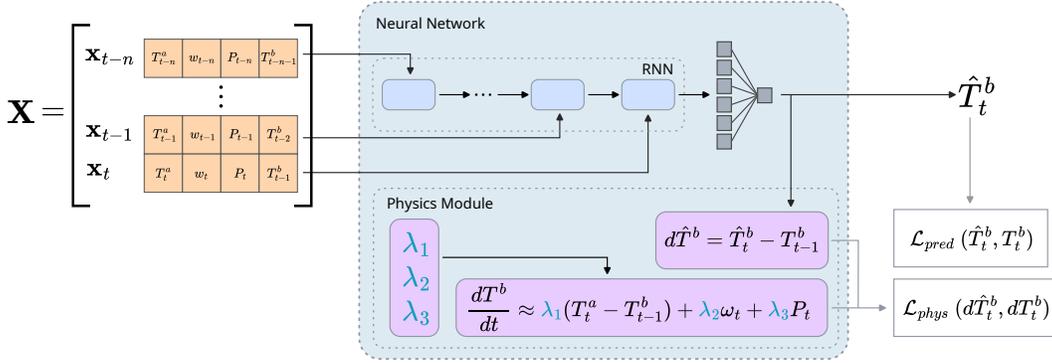


Figure 1: Physics-constrained recurrent neural network (PC-RNN).

single dense layer, reducing the dimensionality of the LSTM output to a scalar value (the temperature prediction \hat{T}_t^b). The detailed network architecture and hyperparameters are described in Appendix B.

For every time step $t - i$, where $i \geq 0$ is the number of lags or previous time steps observed, the network receives the state of the WTG \mathbf{x}_{t-i} as input. Similar to the approach from Zhang (2014), the network also receives information about the current time step t in form of the current state \mathbf{x}_t . The network simultaneously computes the temperature gradient ΔT^b based on Equation 2. Based on Gokhale et al. (2022), a simple Euler-approximation is used to get the predicted temperature gradient:

$$\Delta \hat{T}^b = \hat{T}_t^b - T_{t-1}^b \quad (3)$$

The loss consists of two components, a standard prediction loss \mathcal{L}_{pred} minimizing the error between the temperature prediction \hat{T}_t^b and the true temperature T_t^b and a physics loss \mathcal{L}_{phys} minimizing the error between the neural network computed change in temperature ΔT^b based on Equation 2 and the Euler-approximated temperature change $\Delta \hat{T}^b$ (Equation 3):

$$\mathcal{L} = \mathcal{L}_{pred} + \alpha \mathcal{L}_{phys} \quad \text{with} \quad \mathcal{L}_{pred} = \frac{1}{N} \sum_{i=1}^N (\hat{T}_t^b - T_t^b)^2, \quad \mathcal{L}_{phys} = \frac{1}{N} \sum_{i=1}^N (\Delta \hat{T}^b - \Delta T^b)^2 \quad (4)$$

The physics loss acts as a soft constraint for the PC-RNN to respect the principles of physics embedded in the equation. α is a scaling factor controlling the influence of the physics component on the total loss. The whole setup is shown in Figure 1.

3 EXPERIMENT SETUP

The data used for the experiments consists of three datasets with 10 minutes averaged measurements originating from different wind farms (Plant A, Plant B, Plant C), covering a period from January 2022 to September 2023. See Appendix A for a detailed description of the datasets.

We provide results for a PC-RNN with scaling parameter $\alpha = 0.25$ (fixed a-priori) and two baseline models, a simple RNN using the exact same architecture as the PC-RNN described in Section 2.2 but without the physics component and the following linear model:

$$T_t^b = \theta + \sum_{i=1}^{M+1} \sum_{j=1}^N \Theta_{i,j} \mathbf{x}_{t,i,j} \quad (5)$$

where θ is the intercept, Θ are the model parameters and $\mathbf{X}_t \in \mathbb{R}^{(M+1) \times N}$ is the input data matrix at time step t where the rows $i \in [1, M+1]_{\mathbb{N}}$ are the variable vectors $((T_{t-k}^a - T_{t-k-1}^b), \omega_{t-k}, P_{t-k})$

Table 1: Test results for experiments with a single training WTG.

Model	Plant A RMSE ($\pm \sigma_{\bar{x}}$)	Plant B RMSE ($\pm \sigma_{\bar{x}}$)	Plant C RMSE ($\pm \sigma_{\bar{x}}$)
Linear	0.892 (± 0.041)	1.091 (± 0.031)	1.003 (± 0.034)
RNN	0.667 (± 0.06)	0.652 (± 0.027)	0.844 (± 0.041)
PCRNN	0.636 (± 0.033)	0.74 (± 0.04)	0.803 (± 0.037)

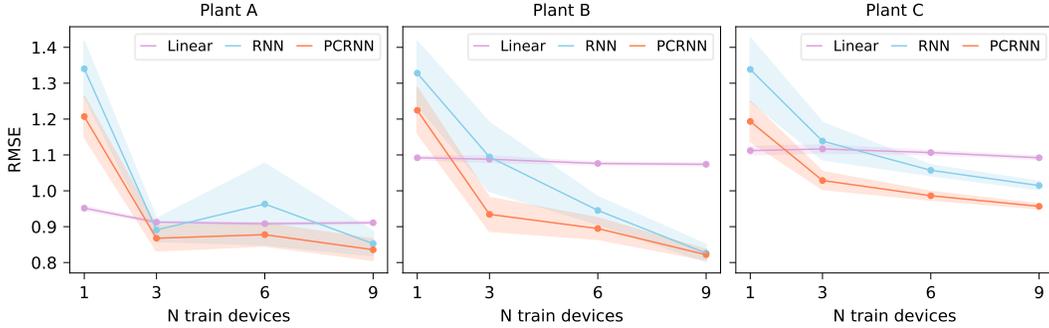


Figure 2: Generalization performance for models trained on Plant B. Error ranges ± 1 SE.

Table 2: Generalization results for experiments with 6 training WTGs.

Train plant	Model	Plant A RMSE ($\pm \sigma_{\bar{x}}$)	Plant B RMSE ($\pm \sigma_{\bar{x}}$)	Plant C RMSE ($\pm \sigma_{\bar{x}}$)
Plant A	Linear	0.887 (± 0.01)	1.078 (± 0.002)	1.101 (± 0.003)
	RNN	0.784 (± 0.017)	1.048 (± 0.048)	1.032 (± 0.015)
	PCRNN	0.726 (± 0.013)	0.95 (± 0.027)	0.968 (± 0.007)
Plant B	Linear	0.908 (± 0.002)	1.076 (± 0.003)	1.107 (± 0.005)
	RNN	0.963 (± 0.113)	0.945 (± 0.038)	1.057 (± 0.014)
	PCRNN	0.878 (± 0.031)	0.895 (± 0.03)	0.986 (± 0.012)
Plant C	Linear	0.982 (± 0.011)	1.118 (± 0.009)	1.059 (± 0.005)
	RNN	0.976 (± 0.049)	1.031 (± 0.024)	0.924 (± 0.016)
	PCRNN	0.929 (± 0.029)	1.028 (± 0.02)	0.913 (± 0.014)

with $k = i - 1$. M is the number of lags or previous time steps given to the model, N is the number of variables. We use $M = 5$ for all models (PC-RNN, RNN, Linear), corresponding to a data interval of one hour. This physics-inspired model is based on Zhang (2014) and Cambron et al. (2017) (see Equation 2), who describe a similar model but without lagged states.

For each experiment, all models are sequentially trained on a subset of every wind farm dataset. We perform experiments with data taken from 1, 3, 6 and 9 randomly sampled turbines to assess performance with different data availability. We perform a time based train test split, using data from the whole year 2022 for training and the data from 2023 as test set for evaluation. Generalization is evaluated with data from 2023 from unseen WTGs in the training plant dataset as well as from the other unseen plant datasets.

4 RESULTS

Test performance. Evaluations on the holdout test sets show very similar performance for the PC-RNN and standard RNN, both outperforming the linear model. When applied as a surrogate model for a single WTG - a common real-world scenario - the PC-RNN offers better performance than the conventional baseline RNN with lower error in two of the three test cases (Table 1).

Generalization. The PC-RNN outperforms both baseline models in a majority of experiments. While the Linear model shows better generalization in experiments with only one training WTG, the PC-RNN performs better in the other scenarios and shows more consistent results than the baseline RNN (Figure 2). Generalization RMSE values for experiments with 6 sampled WTGs for training are shown in Table 2. Complete test and generalization results are shown in Appendix C.

5 CONCLUSION

ML-based predictive maintenance for wind energy systems can help to improve their operational efficiency and reduce costs of wind energy. Integrating physics can increase model performance, but only partially known system dynamics often impede their application. We propose a RNN architecture for predictive maintenance of WTGs constrained by physics where unknown system coefficients are

treated as trainable parameters, allowing to incorporate physics into the model when they are not or only partially known. Using nowcasting of bearing temperatures, this approach is also independent of wind speed forecasts, reducing a large factor of uncertainty. Experiments on different datasets show that the model has state of the art prediction performance and improved generalization performance compared to the baseline models, outperforming them in a majority of tests on unseen environments.

ACKNOWLEDGEMENTS

This work is part of the project DomLearn (892573), which has received funding in the framework of "Energieforschung", a research and technology program of the Klima- und Energiefonds.

REFERENCES

- Philippe Cambron, Antoine Tahan, Christian Masson, and Francis Pelletier. Bearing temperature monitoring of a Wind Turbine using physics-based model. *Journal of Quality in Maintenance Engineering*, 23(4):479–488, 2017.
- James Carroll, Alasdair McDonald, Iain Dinwoodie, David McMillan, Matthew Revie, and Iraklis Lazakis. Availability, operation and maintenance costs of offshore wind turbines with different drive train configurations. 20(2):361–378, 2017.
- Ángel M. Costa, José A. Orosa, Diego Vergara, and Pablo Fernández-Arias. New Tendencies in Wind Energy Operation and Maintenance. *Applied Sciences*, 11(4):1386, 2021.
- Gargya Gokhale, Bert Claessens, and Chris Develder. Physics informed neural networks for control oriented thermal modeling of buildings. *Applied Energy*, 314, 2022.
- Berthold Hahn, Michael Durstewitz, and Kurt Rohrig. Reliability of Wind Turbines. In Joachim Peinke, Peter Schaumann, and Stephan Barth (eds.), *Wind Energy (Proceedings of the Euromech Colloquium)*, pp. 329–332. Springer Berlin Heidelberg, 2007.
- Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.
- George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.
- Alberto Pliego Marugán, Fausto Pedro García Márquez, Jesus María Pinar Perez, and Diego Ruiz-Hernández. A survey of artificial neural network in wind energy systems. *Applied Energy*, 228:1822–1836, 2018.
- M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- M D Reder, E Gonzalez, and J J Melero. Wind Turbine Failures - Tackling current Problems in Failure Data Analysis. *Journal of Physics: Conference Series*, 753:072027, 2016.
- Mahmood Shafiee and John Dalsgaard Sørensen. Maintenance optimization and inspection planning of wind energy assets: Models, methods and strategies. 192:105993, 2019.
- Jinjiang Wang, Yuanyuan Liang, Yinghao Zheng, Robert X. Gao, and Fengli Zhang. An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples. *Renewable Energy*, 145:642–650, 2020.
- Wenxian Yang, Richard Court, and Jiesheng Jiang. Wind turbine condition monitoring by the approach of SCADA data analysis. *Renewable Energy*, 53:365–376, 2013.
- Yigit A. Yucesan and Felipe A. C. Viana. A hybrid physics-informed neural network for main bearing fatigue prognosis under grease quality variation. *Mechanical Systems and Signal Processing*, 171, 2022.
- Zhenyou Zhang. Comparison of Data-driven and Model-based Methodologies of Wind Turbine Fault Detection with SCADA Data. 2014.

A DATASETS

Table 3 shows number of individual WTGs and their maximum power output for each dataset used in the experiments. Distributions of feature values are shown in Figure 3.

Table 3: Number and nominal power of WTGs in each dataset.

	N WTGs	Power [kW]
Plant A	11	850
Plant B	50	850
Plant C	26	850

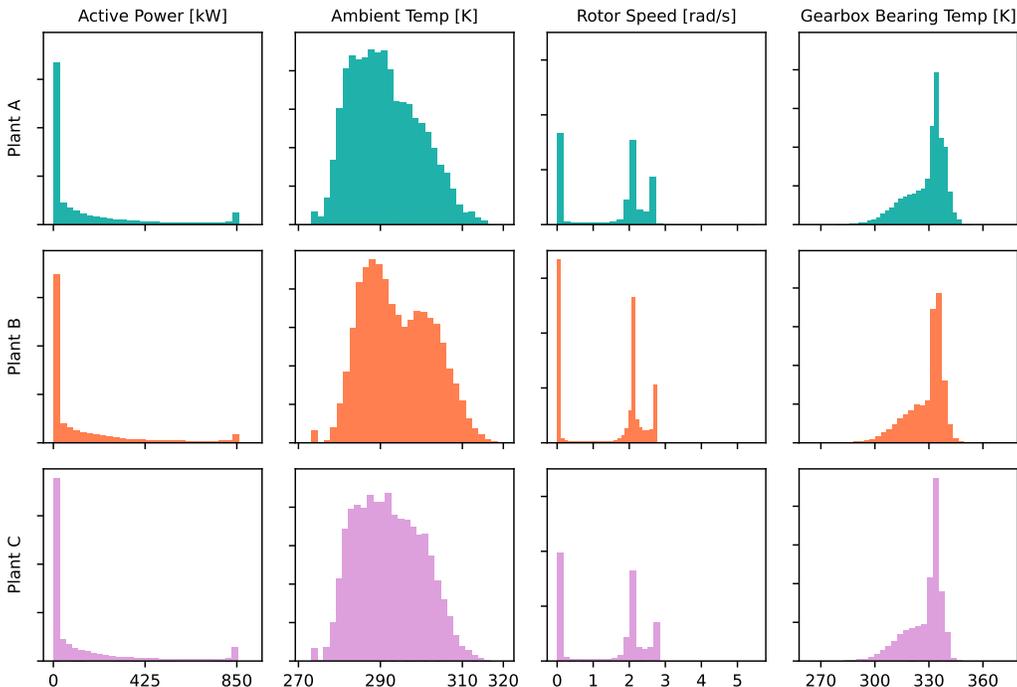


Figure 3: Feature densities for the individual datasets from different wind plants.

B MODEL ARCHITECTURE

Both PC-RNN and baseline RNN consist of a single-layer LSTM cell with a hidden unit size of 16, followed by a dense layer with output size 1. Training was performed using a batch size of 16, a 0.2 validation split and the Adam optimizer with an adaptive learning rate of 0.001. For the PC-RNN, the scaling factor α influencing the impact of the physics loss was set to 0.25. No exhaustive hyperparameter tuning was performed, as this was not the scope of this work.

C RESULTS

Figure 4 shows prediction results and standardized computed temperature changes from one sample PC-RNN. Solving Equation 2 with learned parameters, the model is able to approximate the true pattern of temperature changes, although with smaller magnitude. Complete test results are shown in Table 4.

Generalization performance is shown in Figure 5. In most cases, the PC-RNN shows best performance and lower error variance than the standard RNN. Complete results for generalization experiments are shown in Table 5.

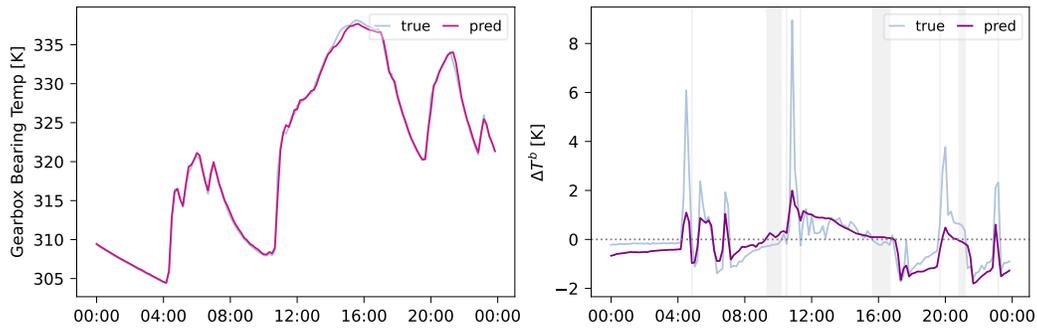


Figure 4: Example of PC-RNN T_t^b prediction (left) and standardized model gradients from the PC-RNN physics component (Equation 2) with learned coefficients (right). Grey areas show time steps where model gradient and true gradient have different sign.

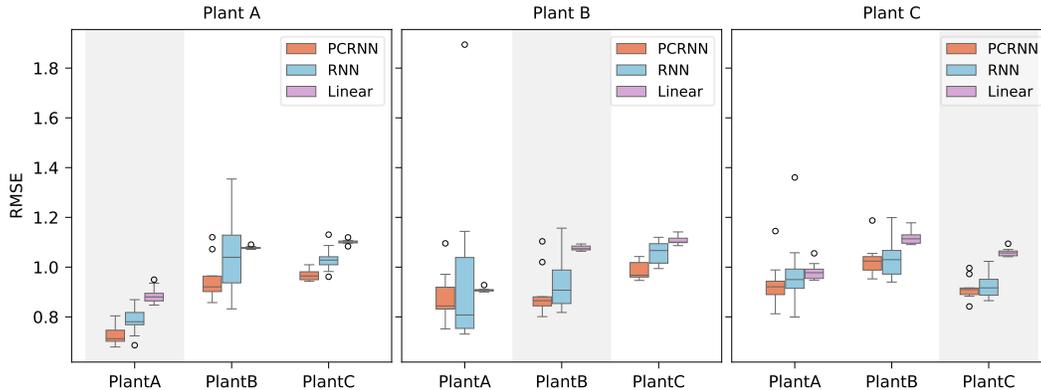


Figure 5: Model generalization performance for training set size of 6 WTGs per dataset. The grey areas mark results for unseen WTGs from training datasets (in-plant generalization).

Table 4: Results on test set for all experiments.

N train	Model	RMSE ($\pm \sigma_{\bar{x}}$)		
		Plant A	Plant B	Plant C
1	Linear	0.892 (± 0.041)	1.091 (± 0.031)	1.003 (± 0.034)
	RNN	0.667 (± 0.06)	0.652 (± 0.027)	0.844 (± 0.041)
	PCRNN	0.636 (± 0.033)	0.74 (± 0.04)	0.803 (± 0.037)
3	Linear	0.914 (± 0.022)	1.013 (± 0.028)	1.021 (± 0.019)
	RNN	0.694 (± 0.047)	0.705 (± 0.035)	0.796 (± 0.025)
	PCRNN	0.679 (± 0.02)	0.722 (± 0.025)	0.808 (± 0.025)
6	Linear	0.914 (± 0.01)	1.063 (± 0.015)	1.012 (± 0.008)
	RNN	0.847 (± 0.045)	0.722 (± 0.017)	0.826 (± 0.02)
	PCRNN	0.768 (± 0.05)	0.777 (± 0.027)	0.867 (± 0.033)
9	Linear	0.904 (± 0.003)	1.056 (± 0.016)	1.026 (± 0.006)
	RNN	0.788 (± 0.028)	0.797 (± 0.085)	0.803 (± 0.008)
	PCRNN	0.772 (± 0.025)	0.741 (± 0.022)	0.839 (± 0.018)

Table 5: Generalization results for all experiments.

N train	Model	Train: Plant A			Train: Plant B		
		Plant A	Plant B	Plant C	Plant A	Plant B	Plant C
1	Linear	0.937 (± 0.005)	1.115 (± 0.007)	1.146 (± 0.012)	0.952 (± 0.005)	1.092 (± 0.004)	1.112 (± 0.011)
	RNN	1.396 (± 0.088)	1.666 (± 0.111)	1.6 (± 0.105)	1.34 (± 0.076)	1.328 (± 0.088)	1.338 (± 0.088)
	PCRN	1.232 (± 0.071)	1.476 (± 0.094)	1.424 (± 0.091)	1.207 (± 0.057)	1.224 (± 0.063)	1.193 (± 0.056)
3	Linear	0.908 (± 0.009)	1.082 (± 0.006)	1.103 (± 0.008)	0.913 (± 0.002)	1.088 (± 0.009)	1.117 (± 0.011)
	RNN	1.015 (± 0.107)	1.29 (± 0.092)	1.136 (± 0.046)	0.891 (± 0.031)	1.094 (± 0.096)	1.139 (± 0.052)
	PCRN	0.855 (± 0.04)	1.04 (± 0.039)	1.013 (± 0.029)	0.868 (± 0.035)	0.934 (± 0.046)	1.029 (± 0.024)
6	Linear	0.887 (± 0.01)	1.078 (± 0.002)	1.101 (± 0.003)	0.908 (± 0.002)	1.076 (± 0.003)	1.107 (± 0.005)
	RNN	0.784 (± 0.017)	1.048 (± 0.048)	1.032 (± 0.015)	0.963 (± 0.113)	0.945 (± 0.038)	1.057 (± 0.014)
	PCRN	0.726 (± 0.013)	0.95 (± 0.027)	0.968 (± 0.007)	0.878 (± 0.031)	0.895 (± 0.03)	0.986 (± 0.012)
9	Linear	0.882 (± 0.014)	1.077 (± 0.001)	1.106 (± 0.001)	0.911 (± 0.004)	1.074 (± 0.003)	1.092 (± 0.004)
	RNN	0.896 (± 0.129)	1.079 (± 0.05)	1.052 (± 0.019)	0.853 (± 0.034)	0.826 (± 0.024)	1.015 (± 0.011)
	PCRN	0.791 (± 0.05)	1.086 (± 0.076)	1.034 (± 0.047)	0.836 (± 0.029)	0.822 (± 0.014)	0.957 (± 0.006)
N train	Model	Train: Plant C					
		Plant A	Plant B	Plant C			
1	Linear	1.006 (± 0.028)	1.141 (± 0.022)	1.1 (± 0.007)			
	RNN	1.576 (± 0.221)	1.58 (± 0.183)	1.421 (± 0.135)			
	PCRN	1.391 (± 0.177)	1.429 (± 0.142)	1.273 (± 0.11)			
3	Linear	1.014 (± 0.019)	1.143 (± 0.017)	1.065 (± 0.005)			
	RNN	1.158 (± 0.072)	1.16 (± 0.054)	1.042 (± 0.042)			
	PCRN	1.082 (± 0.061)	1.135 (± 0.043)	1.003 (± 0.029)			
6	Linear	0.982 (± 0.011)	1.118 (± 0.009)	1.059 (± 0.005)			
	RNN	0.976 (± 0.049)	1.031 (± 0.024)	0.924 (± 0.016)			
	PCRN	0.929 (± 0.029)	1.028 (± 0.02)	0.913 (± 0.014)			
9	Linear	0.986 (± 0.005)	1.119 (± 0.004)	1.055 (± 0.004)			
	RNN	0.998 (± 0.022)	1.031 (± 0.013)	0.956 (± 0.013)			
	PCRN	1.056 (± 0.076)	1.023 (± 0.012)	0.939 (± 0.012)			