

A Data-Driven Approach to Predict NO_x-Emissions of Gas Turbines

Giuseppe Cuccu^{*}, Somayeh Danafar[†], Philippe Cudré-Mauroux^{*},
Martin Gassner[‡], Stefano Bernero[‡] and Krzysztof Kryszczuk[†]

^{*}eXascale Infolab, University of Fribourg, Bd. de Pérolles 90, 1700 Fribourg, Switzerland
[†]Institute for Applied Simulations, ZHAW, Campus Reidbach, 8820 Wädenswil, Switzerland
[‡]General Electric, Brown Boveri Str. 7, 5401 Baden, Switzerland

Abstract—Predicting the state of modern heavy-duty gas turbines for large-scale power generation allows for making informed decisions on their operation and maintenance. Their emission behavior however is coupled to a multitude of operating parameters and to the state and aging of the engine, making the underlying mechanisms very complex to model through physical, first-order approaches. In this paper, we demonstrate that accurate emission models of gas turbines can be derived using machine learning techniques. We present empirical results on a broad range of machine learning algorithms applied to historical data collected from long-term engine operation. A custom data-cleaning pipeline is presented to considerably boost performance. Our best results match the measurement precision of the emission monitoring system, accurately describing the evolution of the engine state and supporting informed decision making for engine adjustment and maintenance scheduling.

I. INTRODUCTION

Modern thermal power plants use gas turbines to generate electricity (example in Figure 1) [1].

Heavy-duty gas turbines are capable of running reliably for extended periods of time requiring little to no maintenance. Over time, however, some engine parts may degrade due to the exposure to thermo-mechanical stress and negatively affect the emission level, in turn enforcing maintenance if the emissions surpass the legal permits.

Such maintenance shutdowns are costly, both due to the operative effort and because of production losses during the intervention, and as such are kept to a minimum. Advanced planning and effective decision support may thereby help reducing the associated costs.

During maintenance, the engine operation set-points are adjusted to a configuration featuring lower emissions without negatively affecting output and efficiency that respect the operative envelope of the gas turbine. A model able to accurately predict engine emissions would enable both to optimally plan maintenance timing, and to prescribe an appropriate adjustment of operation set-points to minimize outage duration and frequency.

Compared to the traditional approaches discussed in Section I-A, this paper explores the feasibility of accurately

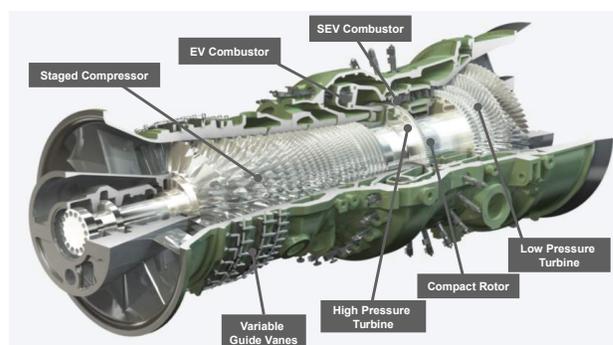


Fig. 1. Architecture and components of a GT24/GT26 gas turbine.

modeling gas turbine NO_x emissions using machine learning, based on historical data collected from long-term engine operation. Sections I-B and I-C highlights the challenges of such an approach and the scientific contributions of the paper. Section II introduces our proposed system, with Section III detailing the most successful configurations. Section III-B presents our best results, achieving a precision on par with sensor-level data. Finally, Section IV offers our conclusions.

A. Monitoring and modeling approaches

Continuous Monitoring and Diagnostics (M&D) of engine performance and health is nowadays standard practice to ensure proper and reliable operation of power plants. Most traditional M&D relies on deterministic systems based on physical rules [2]. This comprises both the investigation of long-term performance trends and the detection of anomalies in order to prevent component failure and avoid costly downtime and repair.

Physical approaches are well suited for modeling degradation and failure mechanisms that are known and can be identified through a limited number of measurements. However, they are more difficult to apply to complex degradation modes

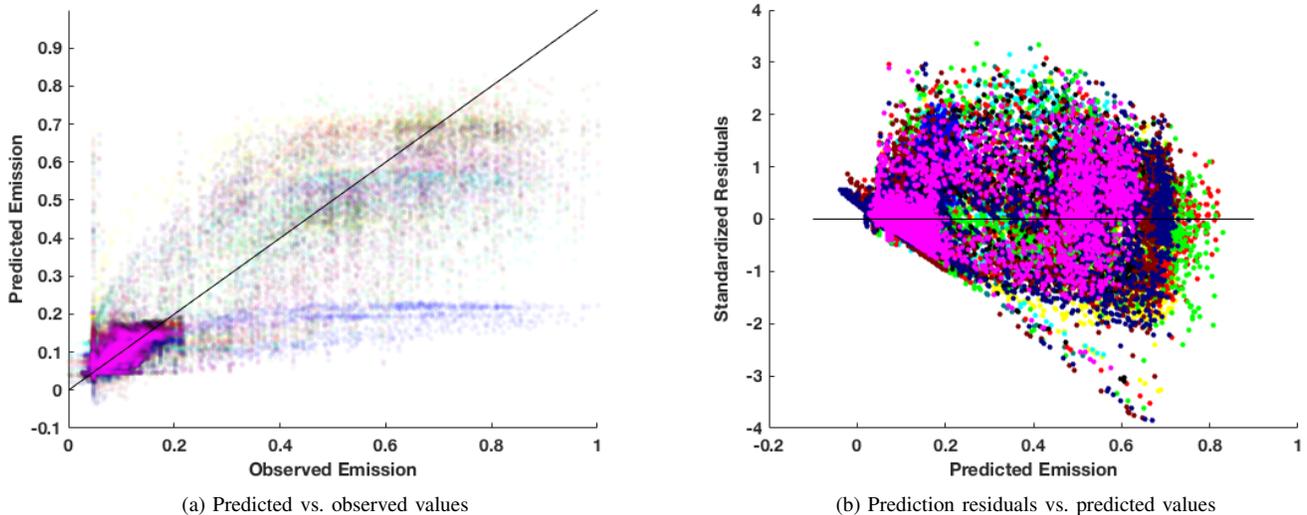


Fig. 2. **Performance on raw (unprepared) data.** Results from one of our best-performing algorithms (ν -SVR) on unprepared (e.g. original, unfiltered) data. The training finds no gradient for improvement: the resulting models have no success on modeling the data.

and evolving engine characteristics where no pattern is easily discernible at first glance.

The application of performance and emission degradation has been addressed in a series of works based on physical modeling approaches. [3] have applied a thermodynamic model of the engine for reconciliation of the commercial operation data to obtain physically consistent information for a precise assessment of degradation. A semi-empirical NO_x emission model has then been identified from engine commissioning data and applied to identify and track degradation at base load [4]. In later work, [5] has consolidated the emission models and extended them to the entire commercial operation range, whereas [6] investigate a similar semi-empirical approach to model thermo-acoustic combustion dynamics. These models provide a detailed understanding of the underlying mechanisms that drive the long-term evolution of combustion behaviour. However, they are relatively cumbersome to set up and maintain. If a detailed physical diagnosis of the behavior is not a primary objective, modern data-based analytics thus bear promise as an alternative, efficient way to enhance combustion monitoring with predictive and prescriptive information.

Such methods have been studied for long in the scientific community, but are still less commonly adopted in industrial practice. An example for gas turbine applications is kernel regression based monitoring of temperature distributions measured at the turbine exhaust, which has shown to be informative on the health status and allows for an early detection of component failure [7].

To our knowledge, applications for emission modeling and monitoring are rare. One example is the investigation of a so-called Predictive Emission Monitoring System (PEMS) based on a neural network, which shall serve as a back-up to cover periods when the measurement-based Continuous Emission Monitoring System (CEMS) is unavailable [8]. The work reported here thus constitutes a scarce investigation of a data-

driven approach for predictive emission monitoring, whose accuracy reaches the precision of the CEMS.

B. Challenges

Working with heavy-duty gas turbines historical data poses a set of unique challenges:

- Working with real historical data is challenging relating to data quality: *sensor noise, sensor biasing, partial observability, observation aliasing, missing*.
- Heavy-duty gas turbines are sophisticated machines working at extreme physical ranges: *high sensitivity to minimal changes, noise amplification, unfeasible (imputed) sensors*.
- Local environment, working conditions and actual usage history make each engine unique: *uniqueness of build, uniqueness of state, local regulations*.

We address these challenges as follows:

- **High noise, bias, imputed readings:** we filter outliers first, and use algorithms robust to noise in the modeling;
- **Observation aliasing, missing data:** we chose models with high generalization capability, and efficient training methods requiring less training data;
- **Data separation:** we selected training algorithms with high performance and low training time, making it feasible to train a different, independent model for each engine.

C. Contributions

This paper introduces the following contributions:

- **Data preparation** (Section II-A): we propose a comprehensive data preparation process that enables the application of a broad range of ML algorithms; selecting only data describing normal operation provides an additional performance boost. Counter-intuitively, better results are achieved through *smaller* training sets, of carefully selected data.

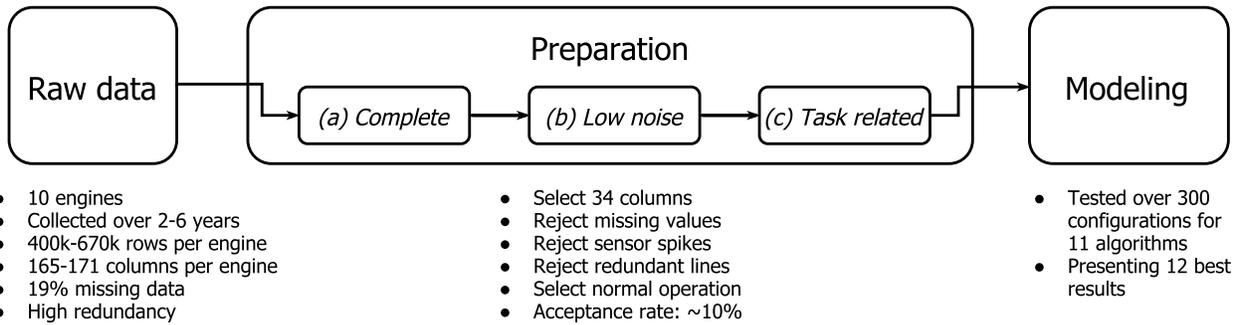


Fig. 3. **System pipeline.** The raw data collected from the engine’s sensors over the years undergoes a tight filtering process: (a) the data is made complete by dropping lines with missing values, (b) noise is lowered by dropping sensor readings outside physical plausibility, and (c) only data pertinent to the task of choice is selected. Only about 10% of the original data is finally used for modeling. The 34 columns mentioned include the 33 model inputs and the target exhaust gas.

- *Comparative study* (Section II-B): we tested a broad selection of machine learning algorithms on our problem. We describe and compare in detail the 12 best performing techniques we tested.
- *High-precision modeling* (Section III-B): our best results achieve a precision comparable to the accuracy of the Continuous Emission Measurement System (CEMS) sensors used in industrial long-term operation.

The next section describes the approach leading to our proposed solution.

II. SYSTEM OVERVIEW

The framework we propose features a machine learning model predicting engine emissions based on the engine configuration and current state (as from sensor readings).

The scope and approach is further detailed in the following sections; an overview is presented in Figure 3.

A. Data preparation

The initially available data amounts to roughly 500’000 rows per engine. Each row holds readings from 180 sensors of various kinds scattered around the engine, such as air pressure after compressor, fuel flow and temperature, and engine vibrations, just to name a few.

Direct application of machine learning techniques to the raw data has proven unsuccessful (Figure 2), highlighting the need for thorough data preparation. The following describes the processing pipeline we put forward in this paper:

- Based on statistical analysis, a subset of the columns is selected to provide predictors with the least redundancy. The dependency of the targets to predictors, and independence of the predictors from each other, are assessed by means of HSIC [9], which takes all possible linear and nonlinear correlation into account.
- Values corresponding to implausible sensor readings (i.e. negative quantities, temperatures below or above real engine ranges, etc.) are discarded as if missing.
- Rows with missing values are simply removed. We considered alternatively using data imputation, but we established

that the remaining, complete data was sufficient to fully train the models.

- Redundant lines (i.e. all values exactly duplicated in another row) are removed. This reduces the training bias towards engine states which are constant in time.
- Since normal engine operation alone is of interest for this study, data recording other engine processes (start up, shutdown, etc.) is carefully selected and discarded as misleading.

The above mentioned process rejects roughly 90% of the original data, culling a mere ~50’000 lines for the modeling phase.

Considering an error margin of 5%, with 95% confidence level and 50% standard deviation of responses, samples of size at least 382 rows are needed for a statistically significant representation of the population (whole data)¹. Own empirical experimentation has confirmed a sample size of 1’000 as ideal for all methods to achieve their best performance, robust to 10% variations.

The data is hence partitioned (no intersection) into 10 training sets of size 1’000, with 1 (common) test set comprising of all the remaining data (~40’000 lines). Since we aim at building a static model (i.e. independent from time), the data constituting each training set is selected randomly (with uniform distribution) across the whole dataset, and then shuffled.

The training sets and the test set are used across all configurations and algorithms without any variation in size, data and order.

B. Algorithm selection

Addressing the limitations found in physically rigorous models (Section I-A), our approach is based on machine learning. In particular, regression analysis addresses predicting a real-valued response $Y = f(X)$ for new values of predictors $X = (X_1, X_2, \dots, X_p)$ [10], [11].

¹Sample size calculator by Raosoft, inc. Available online at: <http://www.raosoft.com/samplesize.html>

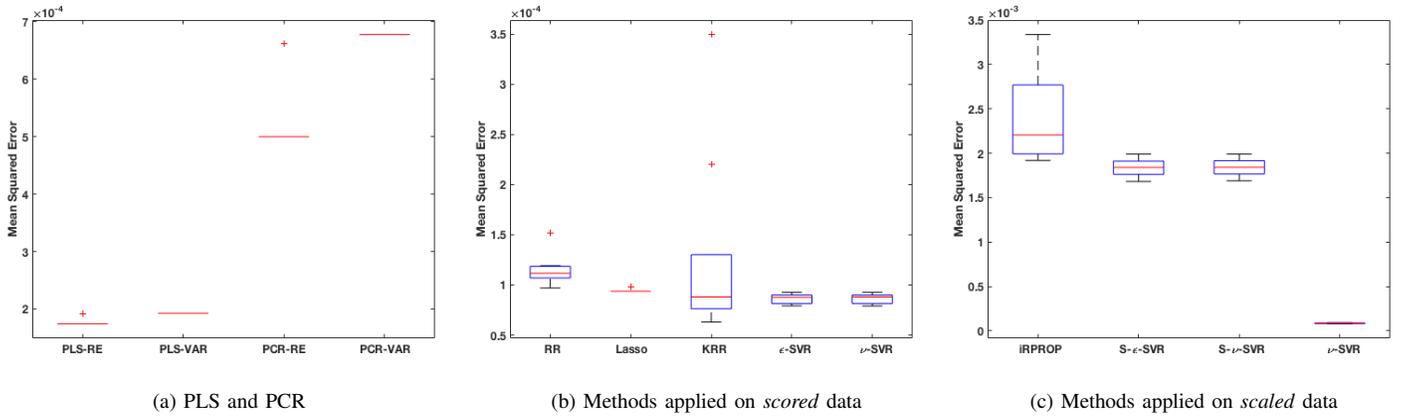


Fig. 4. **Boxplots of Mean Squared Errors.** Please note the scale in plot 4c is one order of magnitude larger: results for ν -SVR from plot 4b are repeated for comparison.

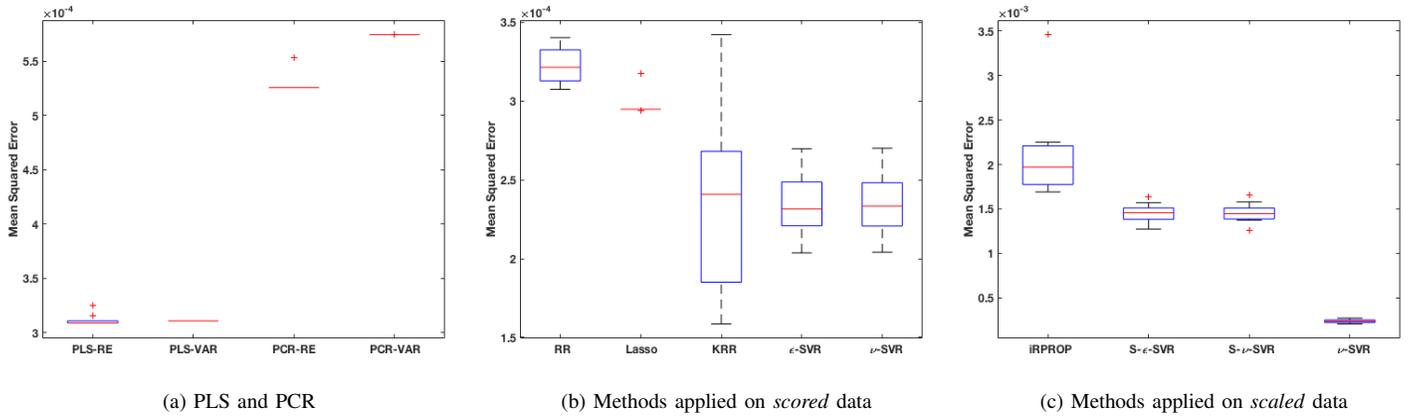


Fig. 5. **Boxplots of Mean Squared Errors,** on data from another engine, for comparison. Please note the scale in plot 5c is one order of magnitude larger: results for ν -SVR from plot 5b are repeated for comparison.

In the context of this project, three main families of algorithms have been considered:

- **Linear regression methods.** Striving for simplicity, and with the goal of establishing a baseline, the following linear regression methods were applied first: *Linear regression* [11], *Ridge Regression* (RR; [12]), *Lasso* [13], *Principal Component Regression* (PCR; [14], [15]), *Partial Least Square Regression* (PLS; [16], [17]).
- **Kernel-based approaches.** Kernelized regression methods are characterized by the usage of kernels. The nonlinear relationship between predictors and outputs is captured in the kernel definition. The following methods were considered: *Kernel Ridge Regression* (KRR; [18]), *Support Vector Regression* (SVR),
- **Feed-forward Artificial Neural Networks** [19]. A parametrized generic function approximator is fitted to predict the current emissions based on the current engine state. The following choice of backpropagation algorithms has been tested: *Incremental Backpropagation* [20], *Batch Backpropagation* [21], *Improved Resilient Backpropagation* (iRPROP; [22]), *QuickProp* [23]

Results shown in III-B come from iRPROP, which in this study consistently achieved best performance. The implementation was based on Fast Artificial Neural Networks (FANN; [24]). Training converged after 20'000 iterations in average.

C. Parameter cross-validation

Method-specific parameters were selected based on 10-fold cross-validation. The following are included for reproducibility:

λ (regularization parameter) in Lasso (10^{-4}), RR (10^{-1}), and KRR (10^{-4}); σ (Gaussian kernel standard deviation) in KRR (10), ϵ -SVR (1) and ν -SVR (1); ϵ (margin) in ϵ -SVR (10^{-2}); ν (number of support vectors) in ν -SVR (0.5); NN: λ (learning rate) = 0.7, η_- (decrease factor) = 0.5, η_+ (increase factor) = 1.2, $\Delta_0 = 0.1$, $\Delta_{min} = 0.0$, $\Delta_{max} = 50.0$, and μ (momentum) = 0, structure $\{in = 33 + b, hid = [10], out = 1\}$ (feed-forward, fully-connected), and activation function $\sigma(x) = \frac{1}{1 + e^{-x}}$.

The number of principal components in PCR-RE was selected by reconstruction error minimization, and in PLS-VAR by variance maximization.

The next section details the proposed system pipeline, and the results obtained by the different configurations.

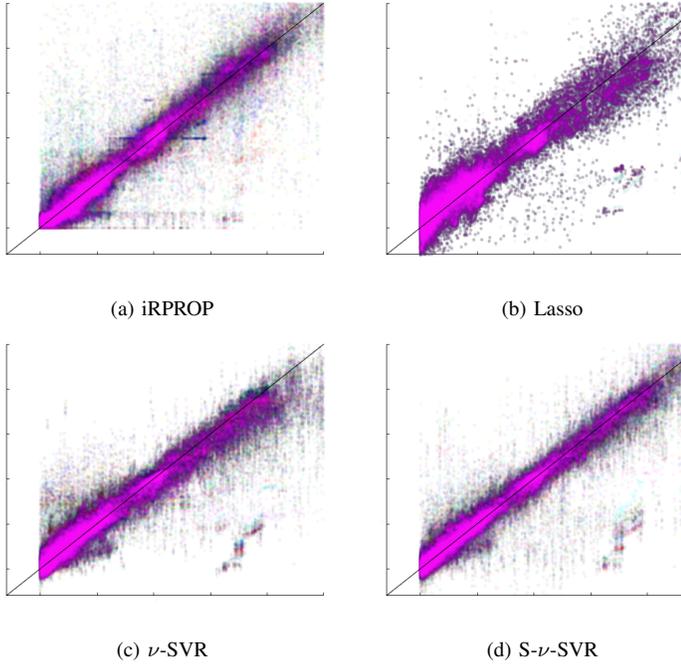


Fig. 6. **Scatterplots of predicted** (vertical axis) **versus observed emissions**. Values on both axis are normalized between 0 and 1.

III. EXPERIMENTAL SETUP

To predict the engine exhaust based on the engine state, each configuration of algorithm plus parameter set is trained on each of the 10 training sets as described in section II-A. The resulting 10 models per configuration are always tested on the common test set, generating 40'000 predictions. Aggregated configuration results correspond to averages over the ten runs.

All regression methods in this study are applied to standardized data (z-score), except iRPROP which works with scaled data (feature scaling). To see the effect of data normalization, we investigate the result of SVR methods with both scaled and standardized data.

A. Performance measures

Standard Mean Squared Error (MSE) of the residuals can become uninformative when the test targets feature high noise. MSE could be improved by overfitting the highest-magnitude outliers, ignoring the model's generalization capability. Section I-B discusses why noise of high magnitude is to be expected in this application.

As a complementary measure of performance, we propose scatterplots of *predicted versus observed emissions*. This allows to intuitively expect clustering along a diagonal line (i.e. $y = x$), with cluster thickness proportional to model precision (plus sensor noise). Partial transparency of the dots proportionally reduces the visual impact of outliers. Please note the different plotting scales in Figure 5c: results for ν -SVR from plot 5b are added for comparison.

Scatterplots of *standardized residuals versus predicted emissions* are also presented: a thick line around $y = 0$ corresponds to a uniform distribution of the predictions with normally

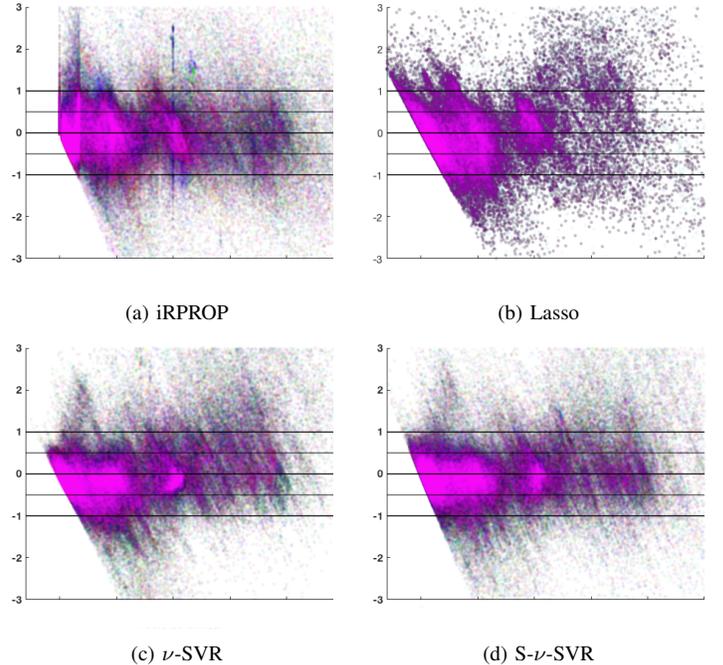


Fig. 7. **Scatterplots of standardized residuals** (vertical axis) **versus predicted emission**. Values of predicted emissions are normalized between 0 and 1.

distributed residuals, as expected of unbiased noise. Any other clear pattern would be indicative of yet-unlearned relations in the data.

For simplicity, results on only one of the available engines are shown unless clearly stated. Performance across engines for the proposed methods remains comparable, as shown in Figure 5.

B. Results

Among linear regression methods, Lasso and RR obtained better results than PCR and PLS (respectively), implying that regularizing the regression function based on 1-norm ℓ_1 or 2-norm ℓ_2 is better suited to the data rather than regularization using PCA. PLS has smaller MSE compared to PCR because it uses the set of principle components that are best suited to predict target variance.

Nonlinear regression methods obtained the best results, accentuating the fact that nonlinear relationships between output and predictors exists. Methods that enforce sparsity worked better than non-sparse methods: for instance, SVR is better than KRR, and Lasso is better than RR. This also highlights the importance of variable selection for this task.

Figure 4 presents a boxplot comparing the MSE of different regression methods over 10 experiments. The MSE is consistently low across methods, in the order of 10^{-4} of scored emissions. Figure 5 shows how comparable results are achieved on other engines.

Figure 6 presents predictions versus observations plots for ν -SVR on both scored (6c, " ν -SVR") and scaled data (6d, " S - ν -SVR"), and Lasso (6b, as the best performing linear method) and neural networks (trained with iRPROP, 6a) on scaled data.

Even though the MSE of ν -SVR is best, the predictions better align with the observations in iRPROP and S- ν -SVR.

The scatterplot of standardized residuals versus predicted emissions in Figure 7 shows the residuals having a narrow range of $[-1, 1]$ in iRPROP and S- ν -SVR, while ν -SVR residuals mostly fall in the $[-1.5, 0.5]$ range. Comparing to ν -SVR, S- ν -SVR and iRPROP, Lasso shows both worse range $[-2, 1.5]$ and worse MSE (Figure 4b).

The algorithms best performing in our study are feed-forward neural networks trained with iRPROP on scaled data (FFNN+iRPROP), and ν support vector regression on scored data (S- ν -SVR). Their MSEs are both one order of magnitude larger than other methods (still minor, at 10^{-3}), but boasting a much tighter clustering around $y = x$ on the predictions vs. observations plot (fig. 6a and 6c). Moreover, the plot of residuals vs. predictions shows normally distributed sensor noise, uniformly spread outliers, and no discernible unlearned patterns (fig. 7a and 7c), implying that the pattern generating the data has been fully learned.

Prediction precision is on the same scale as the expected sensor noise, implying that no better reconstruction can be made without overfitting. Such precision is also deemed sufficient for a live deployment of the application. Both methods train a new model in less than 10 minutes for a single run, easily allowing for daily builds as new data becomes available.

IV. CONCLUSIONS

A model accurately predicting NOx emissions from heavy-duty gas turbines can support informed decisions over engine adjustment and maintenance scheduling. Physically rigorous modeling approaches have found limited application because of the complexity of the involved phenomena and the required effort for model set-up and maintenance in an industrially productive environment.

In this paper, we therefore addressed this problem by applying a selection of machine learning algorithms on carefully prepared data. Our custom filtering pipeline produces complete, low noise, and task-related datasets. We show that our results cannot be reproduced without such thorough pre-processing.

We presented results leveraging a broad collection of machine learning algorithms, ranging from linear regression to neural networks, each assessed on an extensive spectrum of parameter configurations. Comparing the resulting models provides key insight as to the level of sophistication necessary for our application.

Finally, we proposed two methods (FFNN+iRPROP, and S- ν -SVR) capable of reliably training models with the highest feasible precision without overfitting sensor noise. Further analysis shows no unlearned pattern is left in the data. Both methods are comparable in performance and offer short wall-clock running time.

ACKNOWLEDGMENTS

This work was supported in part by the Swiss Commission for Technology and Innovation grant 17900.1 PFIW-IW.

REFERENCES

- [1] F. Güthe, J. Hellat, and P. Flohr, "The Reheat Concept: The Proven Pathway to Ultralow Emissions and High Efficiency and Flexibility," *Journal of Engineering for Gas Turbines and Power*, vol. 131, no. 2, 2008.
- [2] D. Therhorn, "Remote Monitoring and Diagnostic for Combined-Cycle Power Plants," in *ASME Turbo Expo*, 2005, pp. 697–703.
- [3] C. Rudolf, M. Wirsum, M. Gassner, and S. Bernero, "Analysis of Long-Term Gas Turbine Operation With a Model-Based Data Reconciliation Technique," in *ASME Turbo Expo*, 2015, p. V006T05A008.
- [4] C. Rudolf, M. Wirsum, M. Gassner, B. T. Zoller, and S. Bernero, "Modelling of Gas Turbine NOx Emissions Based on Long-Term Operation Data," in *ASME Turbo Expo*, 2016, p. V04BT04A006.
- [5] M. Lipperheide, F. Weidner, M. Wirsum, M. Gassner, and S. Bernero, "Long-term NOx Emission Behavior of Heavy Duty Gas Turbines: an Approach for Model-based Monitoring and Diagnostics," in *ASME Turbo Expo*, 2017, p. V006T05A001.
- [6] F. Weidner, M. Lipperheide, M. Wirsum, M. Gassner, and S. Bernero, "Pulsations in Gas Turbine Operation – Identification and Modeling with the Purpose of Online Engine Monitoring and Optimization," in *ASME Turbo Expo*, 2017, p. V04BT04A011.
- [7] T. Palmé, F. Liard, and D. Therhorn, "Similarity based modeling for turbine exit temperature spread monitoring on gas turbines," in *ASME Turbo Expo*, 2013, p. V004T06A020.
- [8] E. Vanderhaegen, M. Deneve, H. Laget, N. Faniel, and J. Mertens, "Predictive Emissions Monitoring Using a Continuously Updating Neural Network," in *ASME Turbo Expo*, 2010, pp. 769–775.
- [9] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf, "Measuring Statistical Dependence with Hilbert-Schmidt Norms," in *Algorithmic Learning Theory*. Springer, 2005, pp. 63–77.
- [10] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006, vol. 1.
- [11] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1.
- [12] A. N. Tikhonov and V. Y. Arsenin, "Solutions of Ill-Posed Problems," *SIAM Review*, vol. 21, no. 2, pp. 266–267, 1979. [Online]. Available: <http://epubs.siam.org/doi/10.1137/1021044>
- [13] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 58, no. 1, pp. 267–288, 1996.
- [14] H. Hotelling, "The Relations of the Newer Multivariate Statistical Methods to Factor Analysis," *British Journal of Statistical Psychology*, vol. 10, no. 2, pp. 69–79, 1957.
- [15] J. N. R. Jeffers, "Two Case Studies in the Application of Principal Component Analysis," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 16, no. 3, pp. 225–236, 1967.
- [16] H. Wold, *Estimation of principal components and related models by iterative least squares. Multivariate Analysis. Edited by: Krishnaiah PR*. 1966. New York, Academic Press, 1966.
- [17] H. Abdi, "Partial least squares regression and projection on latent structure regression (PLS Regression)," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 1, pp. 97–106, 2010.
- [18] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge: Cambridge University Press, 2004.
- [19] L. Bottou and P. Gallinari, "A framework for the cooperation of learning algorithms," in *Proceedings of the 3rd International Conference on Neural Information Processing Systems*. Morgan Kaufmann Publishers Inc., 1990, pp. 781–788.
- [20] F. Rosenblatt, "Principles of neurodynamics. perceptrons and the theory of brain mechanisms," Cornell Aeronautical Lab inc., Buffalo NY, Tech. Rep., 1961.
- [21] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [22] C. Igel and M. Hüsken, "Improving the Rprop learning algorithm," in *Proceedings of the second international ICSC symposium on neural computation (NC 2000)*, vol. 2000, 2000, pp. 115–121.
- [23] S. E. Fahlman, "Faster learning variations on back-propagation: An empirical study," in *Proceedings of the Connectionist Models Summer School*. Morgan Kaufmann, 1988, pp. 38–51.
- [24] S. Nissen, "Implementation of a fast artificial neural network library (fann)," *Department of Computer Science University of Copenhagen (DIKU)*, 2003. [Online]. Available: <http://fann.sf.net>