Empirical Evaluation of Exponentially Weighted Moving Averages for Simple Linear Thermal Modeling of Permanent Magnet Synchronous Machines

Preprint · March 2019

3 authors:

Wilhelm Kirchgässner
Universität Paderborn

Oliver Wallscheid
Universität Paderborn

Joachim Böcker
Universität Paderborn

Empirical Evaluation of Exponentially Weighted Moving Averages for Simple Linear Thermal Modeling of Permanent Magnet Synchronous Machines

Wilhelm Kirchgässner, Oliver Wallscheid, Joachim Böcker
Department of Power Electronics and Electrical Drives
Paderborn University
33095 Paderborn, Germany
kirchgaessner@lea.uni-paderborn.de

***Abstract*—Permanent magnet synchronous machines (PMSMs) are a popular choice in many traction drive applications due to their high energy and power density and moderate assembly costs. However, electric motor thermal robustness in general is harmed by the lack of accurate temperature monitoring capabilities such that safe operation is ensured through oversized materials at the cost of its effective utilization. Classic thermal modeling is conducted through lumped parameter thermal networks (LPTNs), which help to estimate internal component temperatures rather precisely but also require expertise in choosing model parameters and lack physical interpretability as soon as their degrees of freedom are curtailed in order to meet the real-time requirement. It will be shown that, as an alternative to LPTNS, linear regression achieves similar predictive performance with low computational complexity as long as input representations are preprocessed with exponentially weighted moving averages. Thus, domain knowledge becomes neglectable, and estimation performance depends entirely on collected data and considered input representations. Furthermore, dependence on data quantity and data diversification will be examined in order to assess the minimal mandatory amount of test bench measurements.**

***Index Terms*—Machine Learning, Linear Regression, Thermal Management, Permanent Magnet Synchronous Motor, Temperature Estimation**

I. Introduction

In most automotive applications, a high power and torque density as well as efficiency is required for any given motor, which makes the permanent magnet synchronous machine (PMSM) the preferred choice. In order to exploit the machine's full capabilities, high thermal stress on the machine's potentially failing components must be taken into account. A sensor-based temperature measurement would yield rather precise knowledge regarding the machine's thermal state, yet for the rotor part, it is technically and economically infeasible due to an electric motor's sophisticated internal structure and the difficult accessibility of the rotor. Hence, direct rotor monitoring techniques such as infrared thermography [1], [2] or classic thermocouples with shaft-mounted slip-rings [3] fall short of entering industrial series production. In contrast, stator winding temperature monitoring is measured on a sensor basis nowadays, yet in case of faults, these sensors cannot be replaced due to being firmly embedded in the stator. In addition, sensor functionality may deteriorate during the motor's life cycle. Taking into account the ever increasing importance of functional safety especially in the automotive industry, redundant temperature information becomes obligatory.

Typical approaches for estimating the PMSM's internal components' temperature comprise the setup of electric machine models that provide information about temperature-sensitive electrical model parameters indirectly. There are methods that work with current injection [4] or voltage injection [5] to obtain the stator winding resistance or the magnetization level of the magnets, respectively, as thermal indicators. Moreover, fundamental wave flux observers [6] can help assessing the reversible demagnetization of the embedded magnets. However, these methods suffer from high electric model parameter sensitivity, such that insufficient modeling efforts lead to excessive estimation errors [7].

In this contribution, it will be shown that a simple and well-known machine learning model, namely linear regression, circumvents any kind of electric motor modeling by being fitted on measured test bench data with minor preprocessing directly and still achieves high estimation performance with time-invariant properties. This approach ignores domain knowledge and, hence, does not require any conventional engineering expertise in electric machine thermal management. Estimation accuracy heavily depends on exponentially weighted moving averages (EWMAs) of raw measurement data, which will be motivated and reasoned by analogy to LPTNs in the following. A 60 kW automotive traction PMSM provides the data that denotes the empirical evaluation base.

In order to further boost real-time capabilities of this linear model, trained coefficients will be analyzed, and the importance of different preprocessing parameters will be showcased. This method will be placed among available techniques by evaluating the necessary amount of test bench data against the achieved estimation error.

II. Related Work

In the last decades, related research progress led to approaches that approximate the heat transfer process with equivalent circuit diagrams called lumped-parameter thermal networks (LPTNs). Being partly based on basic formulations of heat transfer theory, they are computationally efficient when reduced to a low-order structure and provide good estimation performance [8]. However, this kind of model must forfeit physical interpretability of its structure and parameter values by significantly curtailing degrees of freedom in favor of the real-time requirement, and, at the same time, expert domain knowledge is mandatory for the correct choice of parameter values. Moreover, the lower the model order of an LPTN is defined, the more empirical data needs to be collected and evaluated for parameter tuning. Therefore, they denote grey-box models, i.e. they are informed by both physical thermal principles and real test bench data.

The first black-box model approach for temperature estimation in PMSMs was introduced in [9]. The models of choice were variations of recurrent neural networks, which represent non-linear and time-varying functions with ten thousands of tunable parameters. A slightly weaker level of estimation accuracy than that of LPTNs' was achieved after having hyperparameters optimized according to an automatic search. Nonetheless, this proves the concept of relying entirely on empirical data when building estimators to be feasible in this domain.

In this work, linear regression models are investigated that, in contrast to (recurrent) neural networks, not only maintain substantially less model parameters to be trained but also hold time-invariant properties contributing to the objective of attaining real-time capabilities.

III. Linear Modeling

Linear regression pertains to a model family of linear approximators that assume a linear relationship between input vectors $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_P)$ and the real-valued (possibly multi-dimensional) output vector $\boldsymbol{y}$ with P being the number of different input representations (e.g. currents, voltages, ambient temperature, etc.). The trainable model coefficients $\boldsymbol{\beta}$ describe the output function $f : \mathbb{R}^P \to \mathbb{R}, \boldsymbol{x}_i \to \hat{y}_i$ applied on observation i [10]:

$$f(\boldsymbol{x}_i) = \beta_0 + \sum_{p=1}^{P} \boldsymbol{x}_{i,p}\beta_p. \tag{1}$$

Estimating $\boldsymbol{\beta}$ from N samples in the training data, one can choose from several methods, whereas the most popular one is the least squares method, which minimizes the residual sum of squares (RSS):

$$RSS(\boldsymbol{\beta}) = \sum_{i=1}^{N} (y_i - f(\boldsymbol{x}_i))^2, \tag{2}$$

$$= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{3}$$

Deriving (3) with respect to $\boldsymbol{\beta}$ and setting it to zero while assuming $\boldsymbol{X}^T\boldsymbol{X}$ is positive definite, yields the following analytically closed solution form for estimations $\hat{\boldsymbol{y}}$:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}. \tag{4}$$

This modeling approach is time-invariant yet not scale-invariant, and standardizing each feature vector $\boldsymbol{x}_p$ is beneficial.

A. Linear Regression With Additional Penalty Terms

The least squares approach or also ordinary least squares (OLS) denotes two drawbacks [10] that become more severe as soon as the feature vectors are not orthonormal, which is the case for most real-world data. Firstly, estimation accuracy suffers from high variance, and, secondly, the scalar model coefficients can not be equated with significance of the corresponding feature as multicollinearity in the data matrix increases variance in $\hat{\boldsymbol{\beta}}$. Adding an additional penalty term to (3) imposes shrinkage gradually stricter to those coefficients that adhere to features with strong multicollinearity. This lowers variance and increases bias in model coefficient determination, albeit it may lead to more solid interpretation of each explanatory variable by evaluating their regression coefficient [10]. In statistics literature, one often finds the ℓ_1 and ℓ_2 norm, which are also referred to as *Lasso* [11] and *Ridge* [12], as possible penalty choices:

$$RSS_{lasso}(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||_1, \tag{5}$$

$$RSS_{ridge}(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda||\boldsymbol{\beta}||_2^2, \tag{6}$$

where λ controls the shrinkage penalty weight. In machine learning literature, the ℓ_2 penalty is also coined as *weight decay*. Both methods, among others, will be probed in order to investigate chances in improved estimation error.

B. Exponentially Weighted Moving Averages

Initially proposed in the econometric field by [13], exponentially weighted moving averages (EWMAs) are a special form of geometric moving averages in which past observations are weighted exponentially decreasing the further in the past they are located. Hence, recent observations contribute more to the current moving average calculation, which, at the same time, is smoothed as there are theoretically infinite past iterations considered as opposed to the classic moving average that will dip and bounce for every extreme point leaving the analyzed window.

The EWMA at time t is computed for a quantity x_t by

$$\mu_t = \frac{\sum_{i=0}^{t} w_i x_{t-i}}{\sum_{i=0}^{t} w_i}, \tag{7}$$

where $w_i = (1-\alpha)^i$ with $\alpha = 2/(s+1)$ and s being the *span* that is to be chosen. The weights of the first s past observations describe approx. 86,5 % of all applied weights. Consecutive calculations of the EWMA can be derived from a more computationally efficient form of (7),

$$\mu_t = (1-\alpha)\mu_{t-1} + \alpha x_t, \tag{8}$$

which is highly relevant especially for automotive applications where embedded systems run on cost-optimized hardware.

C. Analogy to LPTN RC Circuits

It is obvious that LPTNs designed for the investigated use case [8] are characterized by low pass filters or, equivalently, RC circuits smoothing the input quantities and lead to decent temperature estimation accuracy. From signal theory it is known that RC circuits are infite impulse response (IIR) filters of the form

$$x = y + RC\frac{dy}{dt}, \tag{9}$$

which can be discretized to

$$x_t = y_t + RC\frac{y_t - y_{t-1}}{h}, \tag{10}$$

with h being the step size. Rearranging (10) gives

$$y_t = \frac{RC}{RC+h}y_{t-1} + \frac{h}{RC+h}x_t, \tag{11}$$

which resembles (8) with $\alpha = \frac{h}{RC+h}$.

Consequently, it is reasonable to directly apply EWMAs on input quantities in order to obtain regressors exhibiting patterns similar to those in LPTNs. While domain experts would choose the time constant $\tau = RC$ according to machine parameters for LPTN design, here, an automated approach to selecting feasible values for the span s is investigated.

IV. Experiments

As shown by [8], a water-cooled, three-phase PMSM mounted on a test bench yielded the available data, which is consistently sampled at a frequency of $f_s = 2\,\mathrm{Hz}$. In this work, the data comprise the 40-hour investigation basis of [9] but extended by 100 additional hours of measurement with not only low but also especially high fluctuating profiles tracing different random walks in the torque-speed-space that resemble real-world profiles better than the previously available data did. Several independent measurement sequences, each representing a different load profile of variable length with diverse dynamics, were observed. Tab. I compiles the considered quantities that represent the trained models' input and output. All models were trained with the scikit-learn toolbox in Python [14].

A. Cross-Validation

In all experiments, all data representations are standardized on their sample mean and sample unit variance. If not stated differently, five-fold cross-validation (CV) is conducted in every contemplated experiment, and scores are calculated upon the mean squared error (MSE) between predicted sequence and ground truth. During this common CV strategy, the whole dataset is shuffled and split in five equally sized parts, and training is repeated five times with one part denoting the test set and the remaining 80 % being the training set.

TABLE I. Measured input and target parameters

Parameter name	**Symbol**
Inputs	
Ambient temperature	θ_a
Liquid coolant temperature	θ_c
Actual voltage d-axis component	u_d
Actual voltage q-axis component	u_q
Actual current d-axis component	i_d
Actual current q-axis component	i_q
Motor speed	n_{mech}
Derived inputs	
Voltage magnitude $\sqrt{u_d^2 + u_q^2}$	u_s
Current magnitude $\sqrt{i_d^2 + i_q^2}$	i_s
Electric apparent power $1.5 * u_s * i_s$	S_{el}
Targets	
Permanent magnet temperature	θ_{PM}
Stator teeth temperature	θ_{ST}
Stator winding temperatue	θ_{SW}
Stator yoke temperature	θ_{SY}

B. Optimal Span Values

In this experiment, the optimal span values for EWMA feature generation are searched before linear regression is applied. A random search [15] is conducted over an array of possible span values ranging from one minute to two hours with a resolution of one minute. In order to assess the necessary amount of EWMAs per raw feature, one to twelve span values are sampled repeatedly for 30 times each, and five-fold CV is ran with an OLS estimator on each sample. The MSE among all target parameters and folds is optimized in this search. Results, which are similar to those when fitted with Ridge or Lasso, are shown in Fig. 1.

It becomes evident that nine EWMAs from the raw input quantities uniformly distributed along the search space already provide the best regressors in terms of the MSE regardless of the specific choices with no further potential of improvement in case of additional EWMAs. When chosen carefully, three EWMAs are enough to achieve a score of $10.4\,\mathrm{K}^2$, which can be improved only by approx. another $1.5\,\mathrm{K}^2$ by adding more EWMAs.

All following experiments utilize these three optimal span values.

C. Regularizing Costs

Lasso and Ridge models come with a parameter α that controls regularization strength applied on fitted coefficients, where $\alpha = 0$ denotes classic OLS estimation. Different regularization strengths for both models are examined through grid search over α values - see Tab. II. It becomes obvious that applying a constraint on the model coefficients has no or deteriorating effects on model performance. While the Lasso already shows decreased performance at moderate α values, Ridge regression keeps the same score for a wide range of

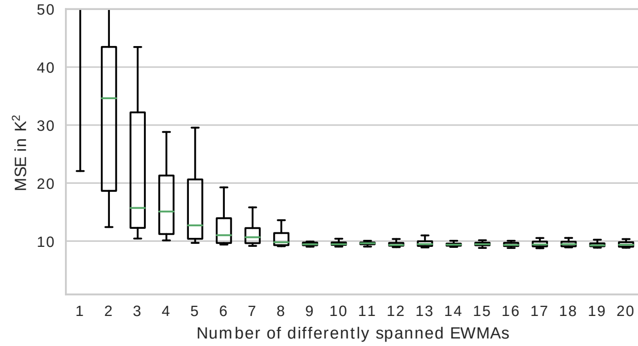

Fig. 1. Random search over an increasing number of differently spanned EWMAs. Each random search involves 30 random samples from a discretized uniform distribution over an array of possible span values ranging from one minute to two hours. Each sample is evaluated by five-fold CV.

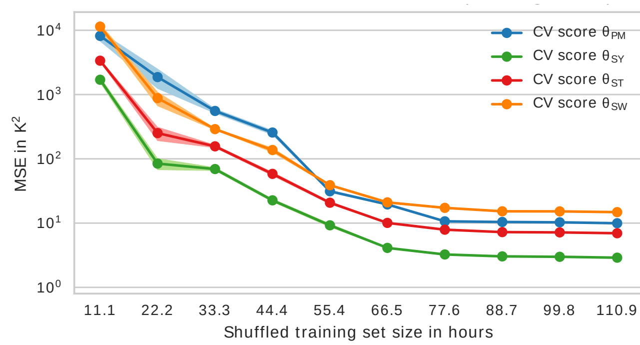

Fig. 2. Varying training set sizes during five-fold CV while keeping the test set size constant at 28 hours. All sets are shuffled. Brighter colors indicate areas within one standard deviation apart from mean. EWMAs are chosen according to optimal values found in the random search experiment.

TABLE II. Grid search over regularization strength α

five-fold CV (Mean over stator and rotor temperatures)[a]			
Lasso		**Ridge**	
α	MSE in K^2	α	MSE in K^2
0.01	13.46	0.1	10.42
0.05	20.05	1	10.42
0.1	29.63	10^2	10.45
0.5	271.54	10^4	11.84
1	442.44	10^6	59.75

[a]OLS MSE: $10.42\,K^2$

weight penalties before it starts degrading rapidly. Hence, adding bias for reduced variance is not beneficial in this application, and OLS regression is pursued in the remainder of this work.

D. Adequate Amounts of Training Data

In order to assess the necessary amount of training data for sufficient estimation accuracy, five-fold CV performance of OLS estimators is visualized over reduced training set sizes while keeping the out-of-fold test set size constant (28 hours). In Fig. 2, it can be seen that test set accuracy does not improve significantly as of roughly 78 hours of measurement data anymore. Thus, assuming the available data covers almost all possible operation points, around 80 hours of records with diverse load profiles can be suggested for training of a robust estimator or at least 67 hours in exchange for a slightly decreased performance.

E. Maximum Test Set Performance

Maximum linear regression performance is leveraged by cross-validating on two load profiles corresponding to seven hours that are unseen during fitting and hold time characteristics similar to those in the remaining data. These two sequences denote the test set on which estimation error is reported in order to assess true generalizability of fitted models, when choosing the optimal span values found earlier. Fig. 3 depicts the predicted load and the ground truth as well as four selected raw input features from the first test profile. In addition, estimation error for each target is delineated with the corresponding maximum deviation L_∞ over the estimation horizon.

It can be inferred from Fig. 2 and Fig. 3 that θ_{SY} is consistently estimated best among the target temperatures, whereas θ_{SW} and θ_{PM} show higher errors when span values are chosen according to the mean score along all four target quantities with equal weights. Moreover, deteriorated accuracy in the beginning and end of the load profile may be due to an unusual operation point and imprecise, longer-spanned EWMA calculations of the first few observations where no past information is available. Note that the permanent magnet temperature is often designed with longer time constants in LPTN implementations and so does its estimation suffer more in the beginning. Further, it can be inferred from preceding experiments that decreased accuracy in the beginning of a load profile is especially severe when the test bench measurement is started with the internal PMSM temperatures not having reached room temperature yet, which highlights strong dependence on sound EWMAs. However, the achieved estimation performance is better than that of an optimized recurrent neural network with 40 hours of training data and a multitude of model parameters in [9].

F. Representation Importance Analysis

A popular take on multicollinearity is quantifying it for each regressor with the variance inflation factor (VIF) and dropping them iteratively until no regressor in the remaining feature set exceeds a predefined threshold anymore [10]. In this approach, one would neglect the dependent variables and estimate each independent variable by fitting on the remaining independent variables. The VIF for a certain regressor is defined as $VIF = (1 - R^2)^{-1}$, with R^2 denoting the coefficient of determination. Feature elimination is conducted in descending order starting with the feature corresponding to the largest VIF. Here, a common heuristic is chosen for the cut-off threshold with $VIF_{cutoff} = 10$, yet estimation accuracy after each elimination is depicted in Fig. 4 to grasp intermediate

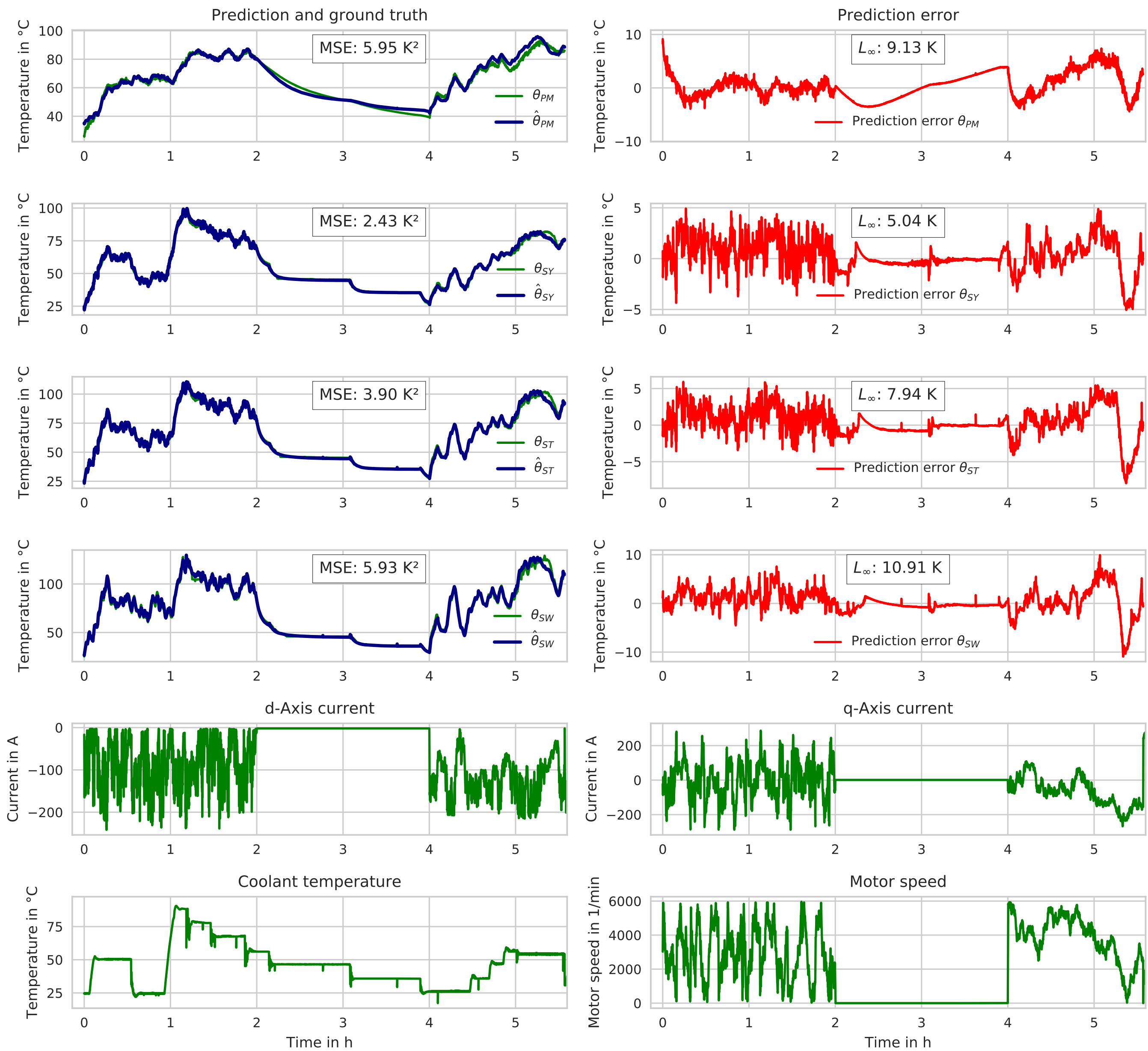

Fig. 3. Test set performance on a 5.5-hour load profile with intermediate cool-down phase. Input representations comprise the three optimally spanned EWMAs found earlier.

performance [16]. It becomes apparent that, generally, performance declines with each input representation eliminated. Interestingly, a significant drop occurs just as of eliminating the 25th most uncorrelated feature (i.e. coolant EWMA with $s = 6\,\text{min}$) in case of the stator temperatures, whereas the rotor temperature estimation accuracy slumps more gradually. Consequently, stator temperatures seem to correlate with the coolant temperature's EWMA to a large extent.

Since weight penalty strategies like Lasso and Ridge diminish performance, feature importance is further determined by evaluating fitted OLS coefficients directly with recursive feature elimination (RFE). Here, OLS on one target temperature is fitted, and the regressor with the smallest model coefficient is dropped iteratively. This is repeated with each target quantity, see Fig. 5. Utilizing the RFE strategy, one can reduce the required feature vectors to roughly 14 input representations per target before performance starts deteriorating slightly. In contrast to the VIF approach, one would have to use different feature sets per target column though.

V. Conclusion and Outlook

In this contribution, it was shown that, for precise estimation of important component temperatures inside PMSMs, one can circumvent the design of white-box models by fitting classic linear regression on benchmark data enriched with differently spanned EWMAs of the original raw features. It was further

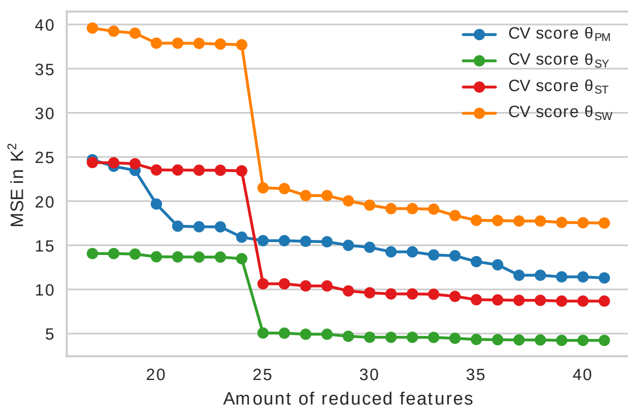

Fig. 4. Five-fold CV with OLS regression on subsets of the original feature matrix. Elimination of features is conducted according to the largest VIF exhibited i.e. independent of target quantitites. Maximum amount of features comprises raw features and the corresponding three best spanned EWMAs while minimum amount is bounded by those that have a $VIF < 10$.

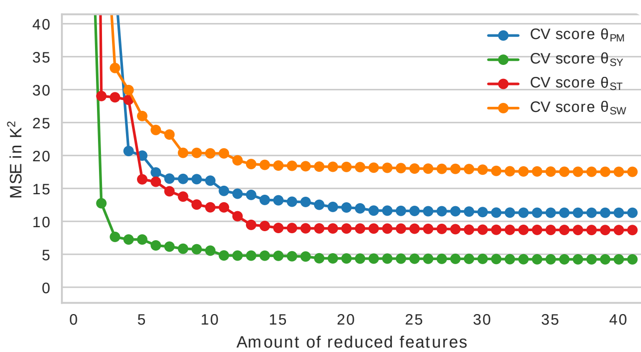

Fig. 5. Five-fold CV with OLS regression on subsets of the original feature matrix. Elimination of features is conducted according to the smallest i.e. most irrelevant coefficients after OLS fitting on each target temperature one after the other. Maximum amount of features comprises raw features and the corresponding three best spanned EWMAs.

demonstrated that the certain choice of span values for EWMA calculation matters less as long as enough span values are sampled from a uniform distribution, although a focused optimization might reveal further chances in model improvement. However, picking the best performing EWMAs from a random search one can achieve high accuracy with as few as three EWMAs per raw feature. Moreover, further reducing the set of features to 30 % - 50 % of the original size is possible with minor performance loss, which makes the presented modeling approach even more feasible for automotive applications on embedded hardware. Real-time capabilities during runtime are not harmed by EWMA calculations due to its computationally efficient form as well as to the slow and steady nature of temperature fluctuations, making inference possible at also low-priority tasks in an embedded system.

Nonetheless, linear regression is an LTI model, yet temperature sequences exhibit strong correlation along neighboring observations, which is not exploited here. Performance gains might be obtained when harnessing non-LTI models together with EWMAs. Future research in this field might aim on answering questions regarding domain adaption (i.e. transfer knowledge of fitted models across motor types), feasibility and accuracy of decision tree based models, and the amount of uncertainty in temperature estimation at runtime.

References

[1] M. Ganchev, B. Kubicek, and H. Kappeler. Rotor Temperature Monitoring System. *The XIX International Conference on Electrical Machines - ICEM 2010*, 2010.

[2] S Stipetic, M Kovacic, Z Hanic, and M Vrazic. Measurement of Excitation Winding Temperature on Synchronous Generator in Rotation Using Infrared Thermography. *IEEE Transactions on Industrial Electronics*, 2012.

[3] C. Mejuto, M. Mueller, M. Shanel, A. Mebarki, M. Reekie, and D. Staton. Improved Synchronous Machine Thermal Modelling. In *2008 18th International Conference on Electrical Machines*, 2008.

[4] S. D. Wilson, P. Stewart, and B. P. Taylor. Methods of Resistance Estimation in Permanent Magnet Synchronous Motors for Real-Time Thermal Management. *Energy Conversion, IEEE Transactions on*, 2010.

[5] M. Ganchev, C. Kral, H. Oberguggenberger, and T. Wolbank. Sensorless Rotor Temperature Estimation of Permanent Magnet Synchronous Motor. In *IECON Proceedings (Industrial Electronics Conference)*, 2011.

[6] A. Specht, O. Wallscheid, and J. Böcker. Determination of Rotor Temperature for an Interior Permanent Magnet Synchronous Machine Using a Precise Flux Observer. In *2014 International Power Electronics Conference (IPEC-Hiroshima 2014 - ECCE ASIA)*, 2014.

[7] O. Wallscheid, T. Huber, W. Peters, and J. Böcker. Real-Time Capable Methods to Determine the Magnet Temperature of Permanent Magnet Synchronous Motors — A Review. In *IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society*, 2014.

[8] O. Wallscheid and J. Böcker. Global Identification of a Low-Order Lumped-Parameter Thermal Network for Permanent Magnet Synchronous Motors. *IEEE Transactions on Energy Conversion*, 2016.

[9] O. Wallscheid, W. Kirchgässner, and J. Böcker. Investigation of Long Short-Term Memory Networks to Temperature Prediction for Permanent Magnet Synchronous Motors. In *Proceedings of the International Joint Conference on Neural Networks*, 2017.

[10] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Verlag, 2nd edition, 2009.

[11] R. Tibshirani. Regression Selection and Shrinkage via the Lasso, 1996.

[12] A. E. Hoerl and R. W. Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 1970.

[13] S. W. Roberts. Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, 1959.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2011.

[15] J. Bergstra and Y. Bengio. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 2012.

[16] R. M. O'Brien. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality and Quantity*, 2007.