Reply to the reviewers' comments for the paper "*Brief communication: Enhanced wind turbine fatigue load estimation using digital shadows with data-driven bias correction.*"

The authors would like to thank the reviewers for their time and for the useful feedback. All the inputs they provided have helped improve the paper.

A list of point-by-point replies to the reviewers' comments is provided below. The reviewer's comments are in black, and our replies are in blue.

A revised version of the manuscript is attached to the present reply, with additions highlighted in blue and deletions marked in red.

Reviewer 1

General comments

The paper has been shortened from the original version, which has impaired replicability. The work is still valid, so only minor revisions are advised. However, the authors should be aware that it will be quite hard for a reader to implement the proposed method based on the concise description presented here.

We thank the reviewer for this constructive comment. We acknowledge that shortening the manuscript resulted in some methodological details being described too concisely. We have therefore revised the manuscript to provide additional information needed to understand and reproduce the proposed approach, including clarification of the bias-correction methodology, the training/testing procedure, and the implementation of the different correction strategies. We have also reorganized related methodological information to provide a more coherent description. We believe these additions substantially improve the reproducibility of the work while preserving the focused scope of the Brief Communication.

Specific comments

- L58: Please clarify what "flapwise" and "edgewise" refer to (displacements, loads, velocities, combinations of those, etc.).

  We thank the reviewer for pointing out this ambiguity. Here, "flapwise" and "edgewise" refer to the flapwise and edgewise components of the static correction force f_0 in Eq. (1b), respectively, rather than to displacements or velocities. We have revised the text accordingly.

- L75: Please clarify how the training vs. testing split is implemented (e.g., random split or time-based split) and the implication for data leak due to the strong autocorrelation of atmospheric and wind turbine time series data.

  We thank the reviewer for this important comment. The training and testing sets were separated temporally rather than obtained by randomly splitting individual samples. In particular, no data from the selected test days were included in the training set. This prevents strongly autocorrelated neighboring samples from being shared between the training and test sets, reducing the risk of data leakage. We have clarified this procedure in the revised manuscript.

- L66 and L90: please unify the description of the BC strategies that are now provided in two separate sections.

  We thank the reviewer for this suggestion. We have consolidated the description of the different BC strategies in Section 2.1, including the shear-based and harmonic-based NN

formulations and the hybrid strategy. The description in the Results section has been streamlined to focus on the training and validation setup, thereby avoiding repetition of methodological details.

- L99: It is unclear why it says that “only the mixed-inflow case is illustrated” while the table right after discusses the other inflow cases. Please remove or clarify.

  We thank the reviewer for pointing out this ambiguity. The statement was intended to indicate that, due to the figure limitation of the brief communication format, only the mixed-inflow cases are shown as time-series figures. The simple- and complex-inflow cases are nevertheless evaluated and reported quantitatively in Table 1. We have revised the text to clarify this distinction.

- L106: Please provide a table for the tower base load too, it is hard to understand which BC correction works better from just one sentence.

  We thank the reviewer for this suggestion. We performed the requested comparison of M_{TB} for all BC strategies. The additional analysis shows that the differences among the corrected strategies are small at the reporting precision used throughout the manuscript and therefore do not provide a meaningful basis for identifying one strategy as superior for the tower-base load. In particular, under complex inflow, the rounded RMSE and DEL errors are essentially identical across the correction strategies.

  This limited sensitivity is consistent with the tower-base load formulation: the steady-state contribution remains unchanged, while the BC primarily affects the perturbation response through the corrected blade states. Consequently, the baseline BC already provides robust estimates of M_{TB}, with only minor changes resulting from the learning-based corrections.

  We have clarified this behavior in the revised manuscript and now report the corresponding M_{TB} performance for the simple, complex, and mixed-inflow cases. Because the strategy-to-strategy differences in M_{TB} are small at the reporting precision used in the manuscript, an additional table would provide little quantitative distinction among the corrected approaches. We therefore explicitly report the corresponding M_{TB} results in the text for the simple, complex, and mixed-inflow cases, while retaining Table 1 for the blade-load quantity, where the correction strategies show clearer differences.

- Fig. 1 and 2: The description of the line colors can be removed, but please explain what the green box is.

  We thank the reviewer for pointing this out. The descriptions of the line colors have been removed from the captions of Figures 1 and 2, and the captions have been revised to clarify that the green dashed boxes highlight periods characterized by variations in wind speed and shear.

Reviewer 2

General comments

Although the proposed learning-based bias correction is potentially useful, I do not consider the dataset and training-validation framework sufficient to establish that it reliably improves fatigue-load estimation across varying inflow conditions

The training data comprise 43 h of simple inflow and 20 h of complex inflow from a single turbine at one site, collected during two campaigns in the same month. The corresponding test sets contain only 6 h and 3 h. High-frequency sampling does not necessarily provide a large number of independent training examples because of temporal correlation. The manuscript does not adequately establish whether these data cover sufficient independent operating conditions for training or how temporal separation between training and testing was ensured. The additional 7.5 h of mixed and transitional inflow provide useful within-turbine holdout validation, but offer limited evidence of performance beyond the experimental setting used for training.

A single-turbine study can provide a valuable proof of concept. However, the present manuscript does not sufficiently establish the adequacy of its training data or the independence and reliability of its validation. Addressing these concerns requires stronger experimental evidence and a clearly documented, temporally separated evaluation, rather than presentation changes alone.

> We thank the reviewer for this important comment. We agree that the original manuscript did not sufficiently document the temporal separation of the training and test datasets and that the number of high-frequency samples alone is not an appropriate measure of the amount of independent information available for training and validation. We have therefore performed an additional dataset audit specifically addressing temporal separation, temporal correlation, the effective number of operating-condition blocks, and coverage of the training and validation conditions. We have also revised the manuscript to frame the present study explicitly as a single-turbine proof of concept, rather than as evidence of transferability across turbines or sites.
>
> First, we verified the training/test split directly from the absolute Date timestamps in each measurement file. The Simple dataset consists of seven training records totaling 43 h and a separate 6 h test record, while the Complex dataset consists of four training records totaling 20 h and a separate 3 h test record. Two additional mixed/transitional records totaling 7.5 h were excluded from training and retained as holdout data.
>
> Importantly, no temporal overlap occurs between any training and test measurement records. For the Simple condition, the minimum separation between the test interval and any training interval is 73 h; for the Complex condition, it is 21 h. Thus, the test data were not constructed by randomly sampling from or selecting neighboring high-frequency samples from the training records. For clarity, the Simple test period lies between earlier and later training records, and we therefore describe the procedure as temporally non-overlapping held-out measurement-record validation rather than chronological forecasting.
>
> Second, we agree with the reviewer that the 10 Hz measurements cannot be interpreted as independent training examples. Our additional autocorrelation analysis confirms substantial temporal dependence in the raw signals. We therefore no longer use the high-frequency sample count as evidence for the effective amount of training information. Instead, we quantified the data using non-overlapping 10-, 30-, and 60-min operating-condition blocks, with blocks constructed separately within each measurement file. At the 10-min scale, the Simple dataset contains 251 training and 35 test blocks, while the Complex dataset contains 116 training and 17 test blocks. The within-turbine

mixed/transitional holdout contains an additional 43 ten-minute blocks. At 30 min, the corresponding training/test counts are 79/11 for Simple and 36/5 for Complex; at 60 min they are 36/5 and 16/2, respectively.

We additionally quantified the residual serial dependence between consecutive block means. For the primary 10-min representation, the 251 Simple training blocks correspond to approximately 41–96 effective blocks, depending on the operating variable analyzed, while the 116 Complex training blocks correspond to approximately 37–60 effective blocks. The sensitivity analysis using 30- and 60-min blocks showed generally lower residual dependence at longer averaging intervals; for example, the effective fractions for the Complex training variables at 30 min are approximately 0.79–0.94. We report these values as approximate dependence diagnostics rather than interpreting individual high-frequency observations as independent realizations.

Third, we evaluated whether the held-out operating conditions are simply repetitions of the training conditions. For the Simple test set, 5.7% of the 10-min blocks lie outside the defined multivariate training-support threshold. For the Complex test set, this fraction is 17.6% at the 10-min scale. Thus, the validation records are temporally separated and include operating-condition combinations that are not fully represented in their respective training datasets.

Finally, the two mixed/transitional periods were kept entirely outside the training procedure. They comprise 43, 13, and 6 complete blocks at 10-, 30-, and 60-min scales, respectively. At the block level, all of these mixed/transitional conditions fall within the multivariate support obtained from the combined training datasets; at the sample level, 4.5% lie outside the defined support threshold. We agree with the reviewer that this remains within-turbine validation and therefore does not establish transferability to other turbines, sites, or campaigns. We have revised the text to explicitly state this limitation.

We believe these additional analyses substantially strengthen the study's experimental basis by demonstrating that the reported results are based on temporally non-overlapping measurement records, explicitly account for temporal dependence, and are evaluated on held-out operating conditions with quantified training-domain coverage. At the same time, we have moderated the manuscript's claims accordingly: the results are presented as a field-based proof of concept for the investigated turbine and operating domain, while validation on additional turbines, sites, seasons, and measurement campaigns remains necessary to establish broader transferability.

The hybrid results also require clarification. Under complex inflow, Table 1 reports a DEL error of 13% for the hybrid strategy, identical to the baseline correction, whereas the harmonic-based NN achieves 4%. Thus, the hybrid provides no improvement over the baseline in this case despite the better performance of the harmonic-based NN. The manuscript should explain why the switching strategy does not capture this improvement. Although this result alone does not disprove robustness, it leaves the claimed benefit insufficiently demonstrated.

We agree that the previous discussion did not sufficiently explain why the hybrid strategy still exhibits a 13% DEL error despite the 4% error achieved by the harmonic-based NN. The hybrid does not select the correction method based on the resulting DEL error; instead, it applies a predefined switching criterion. For this complex-inflow case, that criterion retains the baseline correction and consequently does not exploit the improvement provided by the harmonic-based NN. We have clarified this behavior in the revised manuscript and now explicitly identify it as a limitation of the current switching

strategy. Further refinement of the switching criterion is therefore required to better identify conditions under which NN-based correction provides a clear benefit.

Community's comments

General comments

I reviewed the previous version of this work. The present manuscript is considerably shorter and more focused, and several aspects of the presentation have been clarified. However, I do not find that the principal scientific concern raised previously has been materially resolved. The underlying experimental evidence remains essentially the same, and in my view the manuscript has largely been repackaged and shortened rather than substantially strengthened.

My main concern is the very limited dataset used to establish the performance and generality of the proposed approach. The evaluation is based on measurements from a single 3.5 MW eno126 turbine at one site, with data obtained during two measurement campaigns in October 2020. After filtering, approximately 49 h of simple-inflow data and 23 h of complex-inflow data remain, of which only 6 h and 3 h, respectively, are reserved for testing. An additional 7.5 h of mixed and transitional inflow data are used to examine generalization.

These data may be sufficient to show that the proposed method performs satisfactorily for this particular turbine, site, instrumentation, and measurement period. They are not, in my opinion, sufficient to establish the broader generality of a data-driven bias-correction methodology. The so-called unseen datasets are still obtained from the same turbine, at the same site, using the same instrumentation, and during the same overall measurement campaign. This constitutes holdout validation within a single experimental dataset rather than genuinely independent external validation.

This issue is particularly important for a learning-based correction method. The learned relationship may depend on turbine-specific structural properties, controller behavior, sensor characteristics, atmospheric conditions, and site characteristics. The authors present no evidence that the correction transfers to another turbine, another measurement campaign, another season, or another site. The authors identify wake flow, yawed conditions, offshore conditions, additional turbine platforms, and strongly transient operation as subjects for future work. These are precisely the kinds of conditions needed to establish the method's broader applicability.

A secondary concern is the degree of independence of the load validation. The manuscript states that the digital shadow assimilates blade-root bending moments and provides the training signal for the correction, whereas the blade bending moment at the 25% span location is subsequently used as a validation quantity. A bending moment is an integrated structural response to aerodynamic, inertial, gravitational, and dynamic loading. It is not a unique measure of the underlying aerodynamic load distribution. Furthermore, root and inboard blade bending moments are structurally related quantities. Agreement with the bending moment at another spanwise station is therefore useful as a check on structural-load estimation, but it does not constitute a strongly independent demonstration that the underlying aerodynamic loading has been reconstructed correctly.

The revised manuscript does provide a clearer comparison between the uncorrected digital shadow, the baseline bias correction, the neural-network approaches, and the hybrid method. This is useful. However, these comparisons do not address the more fundamental limitation that all results arise from a relatively restricted experimental dataset. I remain unconvinced that the evidence presented supports the generality the manuscript implies. The work demonstrates the method's performance for the particular turbine and datasets considered, but it has not established broader applicability. In my view, shortening and reorganizing the previous manuscript has not resolved this central limitation.

We agree that the available field measurements originate from a single turbine, site, and measurement campaign and therefore do not constitute external validation across turbine platforms, sites, seasons, or measurement systems. The purpose of the present study is to demonstrate the feasibility and performance of the proposed learning-based bias-correction framework for the investigated turbine and datasets, rather than to claim universal transferability of the trained correction.

To avoid overstating the generality of the results, we have revised the manuscript to clarify the scope of the validation. In particular, the mixed and transitional inflow datasets are described as within-turbine holdout periods that were entirely excluded from training, rather than as independent external validation datasets. These datasets assess performance on held-out operating and inflow conditions while remaining within the same turbine and site dataset. Validation across other turbines, sites, seasons, and measurement campaigns is required before broader transferability can be established, and this is now explicitly stated as an important limitation and direction for future work.

We also agree that the blade-root and 25% span bending moments are structurally related. The blade bending moment at 25% span is not assimilated by the digital shadow and is not used as a training target; therefore, it provides an unassimilated output quantity for evaluating the estimated structural response. However, it should not be interpreted as independent evidence that the complete aerodynamic load distribution has been reconstructed. We have revised the manuscript to make this distinction explicit. The tower-base bending moment provides an additional validation quantity associated with a different structural component and is also not used as an NN training target.

Brief communication: Enhanced wind turbine fatigue load estimation using digital shadows with data-driven bias correction

Hadi Hoghooghi and Carlo L. Bottasso

Wind Energy Institute, Technical University of Munich, 85748 Garching b. München, Germany

Correspondence: Carlo L. Bottasso (carlo.bottasso@tum.de)

Abstract. Building on a previously published digital-shadow framework for wind turbine load estimation, this study augments a linearized aeroelastic model with learning-based bias correction (BC). A neural network (NN) learns operating-condition-dependent corrections while preserving the physics-based model structure. The framework is ~~validated using~~ evaluated using temporally separated field measurements under simple, complex, and mixed inflow conditions ~~. For blade bending moments~~ from a utility-scale wind turbine. For the simple- and complex-inflow test cases, blade bending-moment DEL errors at the 25% span location ~~, damage equivalent load (DEL) errors~~ of 14–24% without correction are reduced to below 5% with the best-performing correction strategies. The largest improvements ~~occur~~ are obtained under complex inflow. A hybrid strategy combines baseline and NN-based corrections ~~, improving robustness across varying operating and inflow conditions.~~ to adapt the applied correction to the prevailing operating conditions. The results demonstrate the potential of learning-based BC to improve fatigue-load estimation within the investigated operating domain.

1 Introduction

Digital twins and digital shadows are increasingly used for wind turbine monitoring, load estimation, and lifecycle management by combining physics-based models with measurement data (Hoghooghi and Bottasso, 2026b, a). In this work, we build on our previously developed digital-shadow framework, which combines an aeroservoelastic model with Kalman filtering to estimate loads (Hoghooghi et al., 2024; Hoghooghi and Bottasso, 2026b). A key challenge is the mismatch between the internal model and the physical turbine arising from model simplifications, parameter uncertainties, unmodelled dynamics, and complex inflow conditions. While ~~bias correction (BC)~~ BC can reduce these discrepancies while preserving the physical model structure (Hoghooghi and Bottasso, 2026b), fixed corrections often have limited effectiveness under changing operating conditions ~~. Our previous work demonstrated the potential of data-driven extensions to improve the adaptability of the digital shadow~~ (Hoghooghi and Bottasso, 2026a).

This study advances that approach by introducing a deep learning-based bias-correction strategy for wind turbine load estimation. An industrial-grade aeroelastic model serves as the internal model of the digital shadow, while an NN learns state-dependent corrections from field data and model predictions. Because the NN augments rather than replaces the physics-based model, the approach retains the ~~physical consistency of the digital shadow~~ physics-based model structure while increasing its flexibility under conditions that the nominal model does not adequately represent.

The main contribution is the development and field validation of an adaptive bias-correction framework for fatigue-relevant load estimation. The proposed method combines conventional BC, which remains effective when the baseline model is reliable, with NN-based correction ~~under~~ to account for more complex inflow conditions. The framework is evaluated using field measurements under simple, complex, and mixed inflow conditions, with a focus on blade-bending-moment DELs. The paper is organized as follows: Sect. 2 describes the methodology, Sect. 3 presents the evaluation, and Sect. 4 summarizes the findings and future work.

2 Methods

The proposed digital shadow builds on the framework presented in Hoghooghi and Bottasso (2026a), combining a linearized aeroelastic model, a Kalman filter, and a learning-based BC module. In contrast to the previous architecture, the optimizer block is replaced by the proposed BC module, which uses measured turbine signals and estimated inflow conditions to improve real-time load estimation.

The digital-shadow formulation, including model scheduling and inflow observers, follows Hoghooghi and Bottasso (2026b). A reduced-order representation of a nonlinear aeroelastic model is obtained by linearization about multiple operating points, with system matrices stored in lookup tables. Rotor-equivalent wind speed and inflow shear are estimated online as described in Hoghooghi et al. (2024); Hoghooghi and Bottasso (2026b, a).

Model–plant mismatch is represented by an additive ~~dynamic~~ condition-dependent static correction $\boldsymbol{f}_0$ and an output bias $\boldsymbol{b}$. The filter model is

$$\dot{\delta\boldsymbol{q}} = \delta\boldsymbol{v}, \tag{1a}$$

$$\dot{\delta\boldsymbol{v}} = -\mathbf{M}^{-1}\left(\mathbf{C}\delta\boldsymbol{v} + \mathbf{K}\delta\boldsymbol{q} + \mathbf{U}\delta\boldsymbol{u} + \boldsymbol{f}_0 + \boldsymbol{\omega}\right), \tag{1b}$$

$$\dot{\boldsymbol{b}} = \boldsymbol{\omega}_b, \tag{1c}$$

$$\delta\boldsymbol{y} = \mathbf{D}_v\delta\boldsymbol{v} + \mathbf{D}_q\delta\boldsymbol{q} + \mathbf{E}\delta\boldsymbol{u} + \boldsymbol{b} + \boldsymbol{\nu}, \tag{1d}$$

$$\delta\boldsymbol{z} = \mathbf{F}_v\delta\boldsymbol{v} + \mathbf{F}_q\delta\boldsymbol{q} + \mathbf{G}\delta\boldsymbol{u}. \tag{1e}$$

Here, $\boldsymbol{q}$ and $\boldsymbol{v}$ denote generalized displacements and velocities, $\boldsymbol{u}$ the inputs, $\boldsymbol{y}$ the measured outputs, and $\boldsymbol{z}$ additional estimated outputs. The matrices $\mathbf{M}$, $\mathbf{C}$, $\mathbf{K}$, and $\mathbf{U}$ represent the mass, damping, stiffness, and control matrices, respectively.

The filter model, including its nine degrees of freedom, input/output definitions, and lookup-table scheduling, is identical to that of Hoghooghi and Bottasso (2026b) and is not repeated here. In the present implementation, the scheduling vector contains only the rotor-equivalent wind speed, $\boldsymbol{s} = \{V\}$. Vertical and horizontal shear and yaw misalignment are estimated online and included through the input vector $\boldsymbol{u}$ using the observer framework described in Hoghooghi et al. (2024); Hoghooghi and Bottasso (2026b, a). In this study, the output bias $\boldsymbol{b}$ is calibrated and subsequently fixed, while the learning-based approach described in Sect. 2.1 determines the ~~dynamic~~ static correction $\boldsymbol{f}_0$ as a function of the operating ~~condition~~and inflow conditions.

2.1 Learning-based bias correction

The two BC terms in Eq. (1), $\boldsymbol{f}_0$ and $\boldsymbol{b}$, are calibrated sequentially to reduce coupling effects. First, $\boldsymbol{b}$ is identified while neglecting $\boldsymbol{f}_0$. The resulting $\boldsymbol{b}$ is then fixed and used to identify $\boldsymbol{f}_0$. Further details are given in Hoghooghi and Bottasso (2026a).

The ~~main contribution of this work is a~~ learning-based strategy ~~for estimating~~ estimates $\boldsymbol{f}_0$ ~~under varying~~ as a function of the operating and inflow conditions. Baseline BC factors are first identified as functions of wind speed and then systematically shifted to generate over- and under-corrected cases, providing a wider range of load-estimation errors for training. The flapwise ~~correction is calibrated before the edgewise correction~~ component of the static correction force f_0 (Eq. 1b) is calibrated first, followed by its edgewise component, because flapwise loading dominates blade fatigue.

The digital shadow is evaluated over the resulting range of BC factors, generating a mapping between operating conditions, correction parameters, and load-estimation errors. A feedforward NN is then trained to learn the inverse relationship and predict correction parameters that minimize these errors.

~~The input space includes wind speed, rotor speed, generator torque, and either~~ The NN therefore estimates BC parameters rather than loads directly.

Two input formulations are considered: a shear-based ~~descriptors (~~formulation using estimated vertical and horizontal shear~~) or~~ , and a harmonic-based formulation using blade-load harmonic components (Hoghooghi and Bottasso, 2026b). ~~Pitch angle may additionally~~ Both formulations additionally include wind speed, rotor speed, generator torque, and blade load, while pitch angle may be included near the Region II/III transition. Separate NNs are trained for simple and complex inflow conditions.

The trained ~~NN provides~~ NNs provide operating-condition-dependent estimates of $\boldsymbol{f}_0$. In the hybrid strategy, the baseline BC is retained unless the vertical- and horizontal-shear indicators exceed 0.4 and 0.05, respectively, in which case the NN-based correction is activated. These thresholds were selected empirically from the available field data and are application-specific.

3 Results

The digital shadow is evaluated using field measurements from a 3.5 MW eno126 wind turbine in northeast Germany (eno energy GmbH). The dataset, collected during two measurement campaigns in October 2020, contains generator torque, rotor speed, pitch angle, tower-top accelerations, and blade-root and tower-base bending moments sampled at 10 Hz. Details on the turbine, instrumentation, and preprocessing are provided by Hoghooghi and Bottasso (2026b).

After removing data gaps, turbine stops, faults, and non-production periods, approximately 49 h of simple inflow and 23 h of complex inflow data remained. Of these, 43 h and 20 h were used for training, ~~leaving~~ while separate 6 h and 3 h ~~for testing~~records were reserved for testing under simple and complex inflow conditions, respectively. The split was performed at the measurement-record level rather than by randomly splitting individual samples. Absolute measurement timestamps confirmed that the training and test records do not overlap, with minimum train–test separations of 73 h and 21 h for simple and complex inflow, respectively. In addition, 7.5 h of mixed and transitional inflow~~conditions,~~ , entirely excluded from training, were ~~reserved for evaluating generalization~~retained as an additional within-turbine holdout. The wind-speed and

shear estimators employed in the analysis were previously developed and validated in Schreiber et al. (2020); Bertelè et al. (2021).

Since the 10 Hz measurements exhibit substantial temporal autocorrelation, individual samples are not treated as statistically independent training examples. Therefore, each measurement record was divided into non-overlapping 10-min blocks, yielding 251 training and 35 test blocks for simple inflow and 116 training and 17 test blocks for complex inflow. Accounting for temporal dependence, these correspond to approximately 41–96 and 37–60 effective training blocks for simple and complex inflow, respectively, depending on the operating variable. Multivariate analysis further shows that 2 of 35 (5.7%) simple and 3 of 17 (17.6%) complex test blocks lie outside the corresponding training-support thresholds, indicating that the held-out data include operating-condition combinations not fully represented during training. The present validation is therefore regarded as a field-based proof of concept within the operating domain of the investigated turbine; transferability to other turbines, sites, and measurement campaigns remains to be established.

3.1 Fatigue-load evaluation

Fatigue performance is assessed using DELs derived from rainfall counting ~~(Natarajan, 2022; Hoghooghi and Bottasso, 2026b) .~~ following Natarajan (2022), with Wöhler exponents of $m = 10$ and $m = 4$ ~~are used~~ for blade and tower-base bending moments, respectively. ~~DELs are computed following the formulation in Natarajan (2022).~~ Prediction error is defined as the relative difference between estimated and measured DELs.

The baseline and learning-based BC approaches are compared for blade-~~root~~bending moments at the 25% span ~~and~~ location and for tower-base bending moments, with an emphasis on the unseen mixed-inflow dataset to evaluate generalization beyond the training conditions.

3.2 Learning-based BC evaluation

~~The learning-based BC is trained using the baseline-corrected factors together with systematically shifted variants derived from field data. The shifts preserve the wind-speed-dependent trend while generating over- and under-corrected cases, thereby providing a range of load-estimation errors for NN training.~~

~~Two input formulations are investigated: a shear-based formulation using estimated vertical and horizontal shear, and a harmonic-based formulation using blade-load harmonics. Both also include estimates of wind speed, rotor speed, generator torque, and blade load. Separate NNs are trained for simple and complex inflow conditions.~~

~~The NN estimates BC parameters rather than loads directly.~~ Blade-root bending moments are assimilated by the digital shadow and provide the training signal for the correction, whereas the blade bending moment at 25% span ($M_{B-25\%}$) and tower-base bending moment (M_{TB}) are used exclusively for validation. Although $M_{B-25\%}$ is not used for training or assimilation, it is structurally related to the assimilated blade-root bending moment and therefore provides validation of an unassimilated structural-load quantity rather than an independent reconstruction of the aerodynamic load distribution.

3.3 Performance under simple and complex inflow

The No-BC, baseline-BC, shear-based NN, harmonic-based NN, and hybrid strategies are compared using independent test data under simple and complex inflow conditions. ~~The learning-based and hybrid BC strategies are evaluated on independent test data under simple and complex inflowconditions. For brevity, only the~~ Due to the figure limitation of the Brief Communication format, only the mixed-inflow ~~case is illustrated~~cases are shown as time-series figures, while the simple- and complex-inflow results are summarized in Table 1.

Table 1 summarizes the root mean squared error (RMSE) and DEL errors for the different correction strategies. Under simple inflow, the baseline BC remains effective, reducing the error from 4% (RMSE) / 14% (DEL) without correction to 3% (RMSE) / 6% (DEL). Under complex inflow, the harmonic-based NN provides the best fatigue-load estimation performance, reducing the DEL error from 13% with the baseline BC to 4%.

Table 1. Overview of RMSE and estimated output DEL errors for $M_{B-25\%}$ under simple and complex inflow conditions.

Inflow conditions	Time duration [hrs]	Estimation error $M_{B-25\%}$ [%] (RMSE / DEL)				
		No BC	Baseline BC	Shear-based inputs	Harmonic-based inputs	Hybrid
Simple inflow	6	4 / ~~15~~ 14	3 / 6	6 / 8	5 / 3	**3 / 4**
Complex inflow	3	3 / 24	3 / 13	2 / 9	**3 / 4**	3 / 13

~~Table 1 summarizes the root mean squared error (RMSE)~~ The hybrid strategy yields RMSE and DEL errors ~~for the different correction strategies. Under simple inflow,~~ of 3% and 13%, respectively, comparable to the baseline BC ~~remains effective, reducing the error from 4% (RMSE) / 15% (DEL) without correction to 3% (RMSE) / 6% (DEL). Under complex inflow conditions, the~~ for this complex-inflow case. Although the harmonic-based ~~NN provides the best fatigue-load estimation performance, reducing the DEL error from 13% with the baseline BC to 4%.~~ NN achieves a lower DEL error, the hybrid strategy does not select the correction retrospectively based on the resulting DEL error; instead, the applied correction is determined by the predefined switching criterion. In this case, the criterion retains the baseline correction and therefore does not capture the improvement provided by the harmonic-based NN. This result highlights a limitation of the current switching strategy and motivates further refinement of the criterion.

For the ~~independent~~ tower-base bending moment, the corrected digital shadow achieves errors of approximately 3% (RMSE) / 2% (DEL) under simple inflow and 2% (RMSE) / ~~6~~5% (DEL) under complex inflow, demonstrating ~~improved estimation of loads that are~~ accurate estimation of a structural load that was not used as ~~NN training targets.~~ an NN training target. Differences in M_{TB} among the individual corrected BC strategies are small at the reporting precision considered here, with no clear advantage of one corrected strategy over another. This limited sensitivity is consistent with the tower-base load formulation, in which the steady-state contribution remains unchanged, while the BC primarily affects the perturbation response through the corrected blade states. Consequently, the baseline BC already provides robust tower-base load estimates, with only minor changes resulting from the learning-based corrections.

Overall, the learning-based correction ~~is most beneficial~~ provides its greatest benefit for blade-load estimation under complex inflow, ~~while~~ whereas the baseline BC remains sufficient under ~~simple~~ the investigated simple-inflow conditions. The superior performance of the harmonic-based NN ~~under complex inflow suggests that~~in the complex-inflow case suggests that, for the investigated conditions, harmonic load signatures provide a more informative representation of inflow complexity than estimated shear alone. Further refinement of the switching criterion is therefore warranted.

3.4 Mixed inflow condition

Two unseen datasets are used to evaluate generalization under mixed and transitional inflow conditions: a 4.5 h post-sunset period with intermittent mixing and a 3 h near-sunrise period with strongly varying wind speed and shear.

For the post-sunset case (Fig. 1), the $M_{B-25\%}$ error decreases from 4% (RMSE) / 15% (DEL) without BC to 3% (RMSE) / 4% (DEL) with the baseline BC. The NN trained on simple inflow achieves 3% (RMSE) / 5% (DEL), while the NN trained on complex inflow further reduces the DEL error to 3%. The hybrid strategy provides the best overall result, with 3% (RMSE) / 3% (DEL). For the ~~independent~~ tower-base bending moment, M_{TB}, the corrected ~~estimates achieve~~ strategies yield approximately 2% ~~(RMSE) / 2% (DEL)~~RMSE and 2–3% DEL errors, with only minor differences among the individual BC strategies.

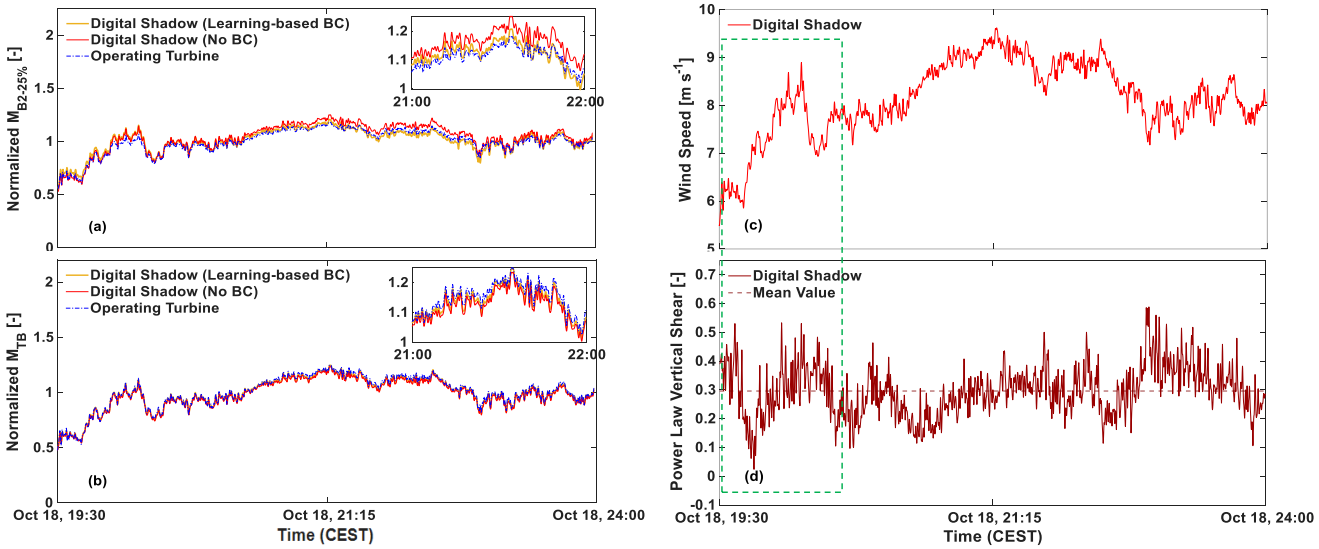


Figure 1. Time histories of normalized $M_{B-25\%}$ **(a)**, M_{TB} **(b)**, wind speed V **(c)**, and vertical shear α **(d)** for 18 Oct 2020 under post-sunset mixed inflow conditions. ~~Measurements are shown as~~ The green dashed ~~blue lines, baseline digital-shadow estimates as solid red lines,~~ box highlights a period characterized by pronounced fluctuations in wind speed and ~~learning-based BC-corrected estimates as solid yellow lines~~vertical shear.

For the near-sunrise case (Fig. 2), the uncorrected digital shadow yields 7% (RMSE) / 6% (DEL) for $M_{B-25\%}$. The baseline BC reduces these errors to 2% (RMSE) / 4% (DEL), whereas the simple-inflow NN gives 5% (RMSE) / 5% (DEL). The hybrid strategy retains the baseline performance at 2% (RMSE) / 4% (DEL), avoiding unnecessary NN corrections when the baseline

BC is already effective. ~~The corrected~~ For M_{TB} ~~estimates achieve 3% (RMSE) / 8% (DEL)~~, the corrected strategies yield approximately 5% RMSE and 8–10% DEL errors, again showing only minor differences among the individual BC strategies.

These unseen cases show that the baseline BC remains highly effective under some mixed-inflow conditions, while the learning-based correction can provide additional improvements when the baseline correction becomes insufficient. The hybrid strategy combines both ~~advantages~~ approaches by retaining the baseline correction ~~when it performs well and activating~~ or activating the NN-based correction ~~when beneficial~~according to the predefined switching criterion.

4

A digital-shadow framework for wind turbine load estimation was extended with a learning-based BC approach ~~that augments an industrial aeroelastic model while preserving its physics-based structure~~to adapt the model correction to operating conditions.

Validation using temporally separated field measurements under simple, complex, and mixed inflow conditions ~~showed that the proposed approach reduced~~ demonstrated improved blade-load estimation. For the simple- and complex-inflow test cases, blade bending-moment DEL errors at the 25% span were reduced from 14–24% without correction to below 5% ~~in the investigated cases~~with the best-performing correction strategies. The largest improvements were obtained under complex inflow, where the harmonic-based NN reduced the DEL error from 13% with the baseline BC to 4%. Under conditions where the baseline BC already provided accurate estimates, the hybrid strategy retained the baseline performance while avoiding unnecessary NN corrections.

The corrected digital shadow also achieved tower-base bending-moment DEL errors of 2–8% ~~, demonstrating improved prediction~~ with the best-performing correction strategies, demonstrating accurate estimation of structural loads that were not used in training. Overall, the results ~~highlight~~ demonstrate the potential of adaptive learning-based BC to improve digital shadow accuracy and fatigue load estimation ~~under variable inflow conditions.~~ within the investigated operating domain. The temporally separated test records and additional mixed/transitional holdout provide a field-based proof of concept, while the present validation remains limited to a single turbine and site.

Future work will ~~extend the framework to~~ assess the broader transferability of the approach by validating it on additional turbine platforms, sites, seasons, and measurement campaigns, including waked, yawed, ~~and offshoreinflow conditions, additional turbine platforms,~~ offshore, and highly transient ~~operating~~ inflow conditions.

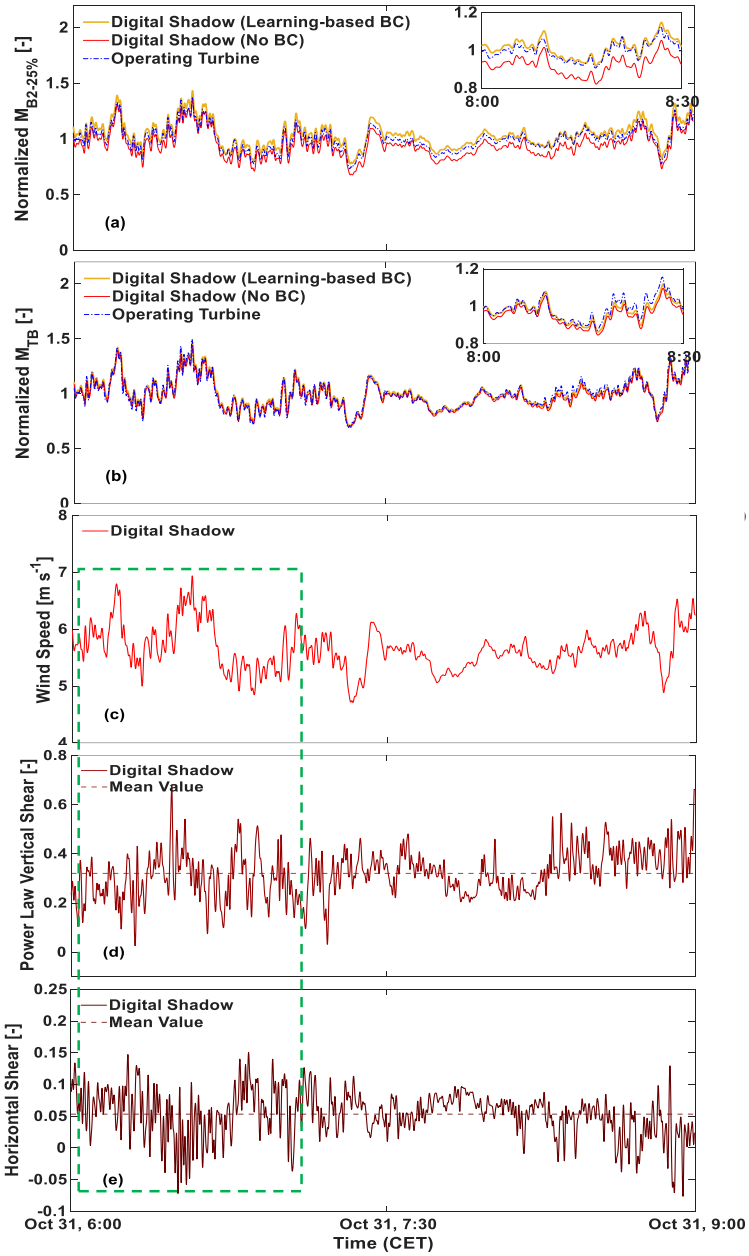


Figure 2. Time histories of normalized $M_{B-25\%}$ **(a)**, M_{TB} **(b)**, wind speed V **(c)**, vertical shear α **(d)**, and horizontal shear k_h **(e)** on 31 Oct 2020 during near-sunrise mixed inflow conditions. ~~Measurements are shown as~~ The green dashed ~~blue lines, baseline digital-shadow estimates as solid red lines,~~ box highlights a period characterized by variations in wind speed and ~~learning-based BC-corrected estimates as solid yellow lines~~in both vertical and horizontal shear.

Appendix A: Nomenclature

$\boldsymbol{b}$	Vector of sensor biases
$\boldsymbol{f}_0$	Static correction force
$\boldsymbol{q}$	Vector of generalized displacements
$\boldsymbol{s}$	Vector of scheduling parameters
$\boldsymbol{u}$	Input vector
$\boldsymbol{v}$	Vector of generalized velocities
$\boldsymbol{y}$	Vector of outputs for Kalman innovation
$\boldsymbol{z}$	Vector of other outputs of interest
$\boldsymbol{\nu}$	Measurement noise vector
$\boldsymbol{\omega}$	Process noise vector
k_h	Horizontal shear
M	Bending moment component
m	Wöhler exponent
V	Wind speed
α	Vertical power-law shear exponent
$(\cdot)^E$	Edgewise component
$(\cdot)^F$	Flapwise component
$(\cdot)^{FA}$	Fore-aft component
$(\cdot)^{SS}$	Side-side component
$(\cdot)_{B-s\%}$	Quantity referred to the $s\%$ spanwise location
$(\cdot)_{TB}$	Quantity referred to the tower base
BC	Bias correction
DEL	Damage-equivalent load
DOF	Degree of freedom
LUT	Look-up table
NN	Neural network
RMSE	Root mean squared error
ROM	Reduced-order model
SCADA	Supervisory control and data acquisition

. All figures and the data used to generate them can be retrieved in Pickle Python and MATLAB formats via https://doi.org/10.5281/zenodo.22207774. The field dataset is the property of eno energy systems GmbH.

. HH developed the new BC training methodology, implemented the digital shadow, performed the numerical simulations, and processed the field dataset. CLB developed the core digital shadow concept and supervised the work. Both authors contributed equally to the interpretation of the results and to the writing of the paper.

. At least one of the (co-)authors is a member of the editorial board of *Wind Energy Science*.

. The technical assistance from Marta Bertelè and Carlo R. Sucameli is acknowledged and greatly appreciated. The authors express their gratitude to eno energy systems GmbH, which granted access to the turbine aeroelastic model and to the field dataset.

. This work has been supported in part by the PowerTracker (FKZ: 03EE2036A), CompactWind II (FKZ: 0325492G), and Life-Odometer (FKZ: 03EE3037B) projects, which receive funding from the German Federal Ministry for Economic Affairs and Climate Action (BMWK). Additional support has been provided by the MERIDIONAL project, which receives funding from the European Union's Horizon Europe program under the grant agreement no. 101084216.

References

Bertelè, M., Bottasso, C. L., and Schreiber, J.: Wind inflow observation from load harmonics: initial steps towards a field validation, Wind Energ. Sci., 6, 759–775, https://doi.org/10.5194/wes-6-759-2021, 2021.

eno energy GmbH: Eno energy GmbH, last access on 15.02.2024.

Hoghooghi, H. and Bottasso, C. L.: Digital shadows with optimized bias correction for accurate wind turbine fatigue load estimation, Journal of Physics: Conference Series, 3224, 062 003, https://doi.org/10.1088/1742-6596/3224/6/062003, 2026a.

Hoghooghi, H. and Bottasso, C. L.: A wind turbine digital shadow for complex inflow conditions, Wind Energy Science, 11, 373–393, https://doi.org/10.5194/wes-11-373-2026, 2026b.

Hoghooghi, H., Bertelè, M., Anand, A., and Bottasso, C. L.: A wind turbine digital shadow with tower and blade degrees of freedom - Preliminary results and comparison with a simple tower fore-aft model, J. Phys.: Conf. Ser., 2767, 032 026, https://doi.org/10.1088/1742-6596/2767/3/032026, 2024.

Natarajan, A.: Damage equivalent load synthesis and stochastic extrapolation for fatigue life validation, Wind Energ. Sci., 7, 1171–1181, https://doi.org/10.5194/wes-7-1171-2022, 2022.

Schreiber, J., Bottasso, C. L., and Bertelè, M.: Field testing of a local wind inflow estimator and wake detector, Wind Energ. Sci., 5, 867–884, https://doi.org/10.5194/wes-5-867-2020, 2020.